Comparative Study of Machine Learning Model For Smart Solid Waste Management

Nikhil Kalme
School of Computer Engineering
MIT Academy Of Engineering
Alandi, Pune, India
nikhil.kalme@mitaoe.ac.in

Mohammed Ismail
School of Computer Engineering
MIT Academy Of Engineering
Alandi, Pune, India
mohammedismail@mitaoe.ac.in

Ashwini Kekan
School of Computer Engineering
MIT Academy Of Engineering
Alandi, Pune, India
ashwini.kekan@mitaoe.ac.in

Apeksha Patil
School of Computer Engineering
MIT Academy Of Engineering
Alandi, Pune, India
apeksha.patil@mitaoe.ac.in

Prachi Jadhav
School of Computer Engineering
MIT Academy Of Engineering
Alandi, Pune, India
prachi.jadhav@mitaoe.ac.in

Sharmila Kharat
School of Computer Engineering
MIT Academy Of Engineering
Alandi, Pune, India
sharmila.kharat@mitaoe.ac.in

***Abstract*—Object detection, a pivotal task in computer vision, involves identifying and locating objects in images. Modern deep learning models, including YOLOv5, RetinaNet, and CNNs, have excelled in this domain, demonstrating state-of-the-art results on benchmark datasets. YOLOv5 stands out as a single-stage model renowned for its speed and accuracy, utilizing a singular neural network to predict bounding boxes and class labels. In contrast, RetinaNet and CNNs operate as two-stage models, employing a region proposal network and a subsequent classifier for object detection. This paper conducts a comparative study evaluating the image detection accuracy of YOLOv5, RetinaNet, and CNNs on a custom dataset featuring over 2000 images across three object categories: plastic bottles, glass, and metal cans. Results showcase YOLOv5's superior computational speed and image detection accuracy, achieving a mean average precision (mAP) of 50.7% on the COCO dataset, surpassing CNNs with a mAP of 43.2%. Not only does YOLOv5 exhibit heightened accuracy, but it also processes images faster, requiring only 13 milliseconds on a GPU compared to CNNs' 32 milliseconds. These findings position YOLOv5 as an optimal choice for real-time object detection applications, making it superior in accuracy and speed for diverse applications such as autonomous driving, robotics, and video surveillance.**

***Index Terms*—Object Detection, YOLOv5,CNN**

I. Introduction

The effective detection and management of solid waste are essential for mitigating environmental pollution and ensuring a sustainable future. The rapid growth of urban populations and industrial activities has led to an unprecedented increase in the generation of solid waste, posing significant challenges for waste management systems worldwide. To address this issue, advancements in computer vision and machine learning techniques have paved the way for more efficient and automated solid waste detection.This paper presents a comparative study between YOLOv5 (You Only Look Once version 5) and traditional Convolutional Neural Networks (CNNs) for the purpose of solid waste detection. YOLOv5 is a state-of-the-art object detection model known for its speed and accuracy, while CNNs have been widely used for image analysis tasks. [1] The objective of this study is to assess and compare the performance of these two approaches in the context of solid waste detection, with the aim of determining which one offers superior results in terms of accuracy and efficiency.The remainder of this paper is organized as follows: Section II provides an overview of related work in the field of waste detection using computer vision. Section III details the methodology and datasets used in our comparative analysis. In Section IV, we present the experimental results and discuss the findings, and in Section V, we conclude the paper with implications and potential future work.By conducting this comparative study, we aim to contribute to the development of more effective and automated solid waste detection systems, which can significantly enhance waste management practices and contribute to a cleaner and more sustainable environment.

II. Related Work

There are different kinds of object detection algorithms; some are more conventional than others, having been created relatively recently. These architectures vary from one another in terms of the hardware resources needed, accuracy, and speed. With the development of deep learning, neural networks, and the availability of large-scale image datasets, significant progress has been made in the field of object detection throughout time. In order to automate waste management in outdoor contexts, a study uses ”SWDet” introduces an anchor-based object identification framework that is specifically intended for solid trash detection in aerial photography. This model uses a new focal loss function that makes it possible to identify waste objects more effectively than with traditional techniques. SWDet improves waste detection in aerial images, creating new opportunities for solid waste management automation that are advantageous to the environment and municipalities alike. [1]

The another study presents "Tiny SSD," a small-sized deep convolutional neural network for Single-Shot identification (SSD) that is designed for embedded and real-time object identification applications. Tiny SSD minimizes processing resources while providing competitive detection performance, addressing the trade-off between efficiency and accuracy. The network architecture is appropriate for real-time inference since it combines bounding box prediction and feature extraction in a single pass. Tiny SSD achieves a compact size without sacrificing detection accuracy by using a carefully thought-out model design and quantization process. This research highlights how it may be applied in situations with limited resources, such embedded systems and edge devices, indicating that it is a viable option for effective object detection. [2] Another study presents technique for identifying objects in noisy images .Using Single Shot MultiBox Detector (SSD), the issue with noisy image identification has been found and examined.It focuses on overcoming the difficulties presented by noisy environments, where conventional methods of object detection might not work well. To improve object recognition accuracy, the suggested method combines cutting-edge object detection algorithms with denoising techniques. In practical applications, the technique enhances object detection performance by efficiently cutting down on image noise. [3]

One study offers a novel method for object detection in situations with a dearth of labeled data. They present a framework for few-shot object detection that makes use of semi-supervised learning by training on both labeled and unlabeled data. To annotate unlabeled data, pseudo-labeling is used, and consistency regularization makes sure the model's predictions hold true for various views of the same image. To compare the EPIC-KITCHENS-55 dataset's performance,they employed FCOS as an anchor-less detector and Faster R-CNN and Cascade R-CNN as anchor-based detectors. Backbones were pre-trained with ImageNet using ResNet-50, ResNet-101, ResNeXt-101, and HRNetV2p-W32; training details were used for each combination of backbone and head structure. Most likely, the paper addresses popular few-shot learning strategies and provides experimental support to show how successful their method is. [4] The another study includes an innovative approach to address the intricate task of detecting diminutive objects flying at low altitudes. It leverages a specially designed lightweight Convolutional Neural Network (CNN) that is enriched with advanced feature extraction capabilities. This method directly tackles the challenges posed by the detection of small, near-ground objects, a facet often marginalized in conventional object detection methodologies. The enhanced feature extraction empowers the model to significantly augment its performance in the detection of these small objects. Furthermore, the network's lightweight architecture ensures computational efficiency, rendering it ideal for real-time applications. [5]

An another study presents extensive overview of deep learning-based generic object detection architectures, tactics, applications, and current. It offers a thorough examination of the many methods and strategies applied to object detection, emphasizing the development of deep learning models. It explores the wide range of techniques and approaches, including RetinaNet, Feature Pyramid Networks (FPN), Region Proposal Networks (RPN), MobileNets, EfficientDet, CenterNet, Non-Maximum Suppression (NMS), Single Shot MultiBox Detector (SSD), You Only Look Once (YOLO), and anchor-based and anchor-free detection methods.It also covers the latest developments in this area, such as the application of one-shot learning, transfer learning, and the fusion of object detection with other computer vision tasks. [6]

A survey based study offers a thorough overview of datasets and object detection techniques based on deep learning that are specifically designed for overhead imagery analysis. In this study, various state-of-the-art methods are examined and their architectures and application domains are presented. These methods include SSD, YOLO, R-CNN, and Faster R-CNN. Additionally, it highlights the value of datasets in the assessment of object detection models by presenting a range of datasets appropriate for data captured by drones, satellite imagery, and aerial photos. Through the consolidation of these insights, this survey facilitates the development of cutting-edge applications for overhead imagery analysis and is an invaluable resource for researchers and practitioners working in the fields of computer vision and remote sensing. [7]

A. *Literature Gap and Proposed Solution*

Several research have been conducted on the use of various deep learning techniques for object detection, according to the literature review that was provided. Nevertheless, a number of research gaps remain that require attention despite the advancements made in this field. One of the gaps is the lack of comparative study on various object detection methods. The efficiency of object detection strategies in increasing accuracy has been the subject of numerous studies; however, in order to determine which strategy performs best for object detection, it is necessary to compare its performance with that of other models. Even though numerous studies have shown encouraging results in terms of increasing the accuracy of object detection, more research is required to determine how these techniques can be used in real-world scenarios. This leads to taking into account elements like cost-effectiveness and scalability when assessing the feasibility of these methods. Ultimately, the review of the literature shows that, although object detection accuracy is crucial, there are other critical conditions that should also be taken into consideration, such as the ability to detect objects in noisy images. Therefore, further study is required to fully understand how deep learning techniques can be applied to object detection.

III. Proposed Methodology

In this research study, a Convolutional Neural Network (CNN) was created for the purpose of feature extraction and classification within the training dataset. The choice of a 3x3 convolutional filter size was made due to its suitability for handling smaller objects. The network incorporated ReLU activation functions in various hidden layers and utilized the categorical cross-entropy loss function in conjunction with the Adam optimizer set at a learning rate of 0.001. The architecture of the network, depicted in Figure 1, illustrates the presence of distinct hidden layers, filters, and pooling operations that were applied to the neural networks.

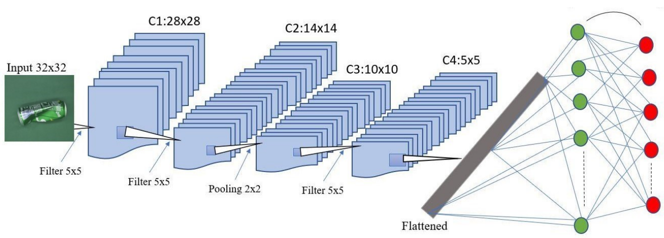

fig.1 Convolutional Neural Network Used for classification.

A. Preprocessing

Before this study gets into the detailed description of the applied methodologies, it is advantageous to provide context regarding the dataset as well as certain design choices and challenges encountered during the initial phases of working with the dataset. This particular dataset, which pertains to solid waste, encompasses a collection of more than 2000 images that have been annotated for the purpose of classifying materials as ”Metal,” ”Plastic,” and ”Glass.” [9] Notably, a noteworthy observation was that images depicting metal objects would frequently contain multiple instances of metal within the image frame.

The labels were modified to provide clear and distinct identification between ”Glass” and ”Plastic.” Additional challenges encountered with this dataset included the visual similarities between the images of glass and plastic materials. As the process of manual annotation and bounding box delineation advanced, it became progressively difficult to discern the discrepancies between the two categories. [9]

The dataset was divided into three distinct subsets: a training set, a validation set, and a testing set, each intended for utilization with YOLOv5. With a total of 2,000 annotated images, these annotations were evenly distributed, resulting in 800 annotations for each classification category. This even distribution of annotations played a crucial role in enhancing the training performance of YOLOv5.

During the classification phase, a series of image transformations were applied. Initially, the images were resized to a uniform size of 32x32 pixels. Following this resizing step, the images were converted to grayscale. This conversion to grayscale was carried out due to the understanding that the color information within the images did not significantly contribute to the classification process. Moreover, it served to simplify the processing complexity of the Convolutional Neural Network (CNN). The resizing step was essential because the original dataset exhibited a wide range of image sizes, ranging from 15x15 to 250x250 pixels. Standardizing the image size was necessary as the CNN requires input images to be of consistent dimensions before they are fed into the network.

B. Training an Object Detection Model

To facilitate model training on the dataset, the cloud-based platform Google Colab was employed. Google Colab provides the advantage of free GPU runtime and operates within a Python-based Jupyter notebook environment. Within this environment, it is possible to clone the public GitHub repository for YOLOv5 and perform the installation of all requisite Python packages.

Python, the preferred programming language of data science engineers, comes equipped with an array of tools for converting outcomes into informative graphs. Additionally, it enables the deployment of the trained YOLOv5 model to generate images featuring classification bounding boxes and associated confidence levels.

In the initial training attempt of a YOLOv5 model, we supplied the following arguments to the train.py script:
python train.py –img 416 –batch 10 –epochs 150 –data dataset.location/data.yaml –weights yolov5s.pt –cache

The ”img” flag specifies the dimensions, in pixels, of the input images, both in terms of length and width. Meanwhile, the ”batch” flag governs the batch size, indicating how much data can be loaded into memory at once. This value is contingent on the hardware specifications in use.

A total batch size of 10 was chosen as the recommended default, although it’s worth noting that the YOLOv5 documentation advises against using very small batch sizes, as they can lead to suboptimal batch normalization statistics. The ”epochs” flag corresponds to the number of complete passes that the training dataset undergoes during training, with 150 being the suggested default value. Adjusting the number of epochs is a critical parameter to fine-tune as it helps in identifying potential overfitting or underfitting issues. Overfitting can have a detrimental impact on the model’s performance and its ability to generalize to new data. The ”data” flag specifies the file location of the training data, while the ”weights” flag points to the YOLOv5s.pt checkpoint, a smaller COCO pre-trained model. This choice takes into account the time limitations of the cloud-based training environment, which allows for free usage for less than 24 hours.

C. Post-Processing

In our first attemot to train a customized YOLOv5 model, we followed certain parameter recommendations for the train.py script. Upon reviewing some of the visual charts generated by the Convolutional Neural Network (CNN), it became apparent that our model was experiencing overfitting. This was evident from an undesirable positive slope observed towards the end of the bounding box regression loss graph or box loss graph. To address the overfitting issue, the number of training epochs was subsequently reduced from 150 to 120. This adjustment aimed to limit the frequency with which the training data was processed through the algorithm, thereby mitigating the overfitting problem.

IV. EXPERIMENTAL RESULTS ANALYSIS

From the fig.2, YOLOv5 and retina net have better accuracy than the other two models, i.e. CNN and SSD.

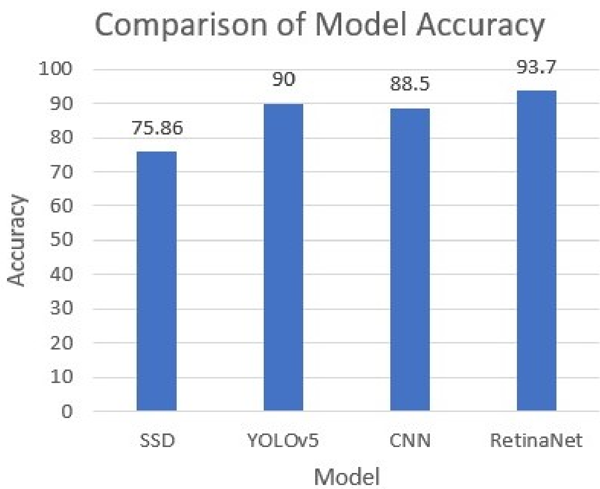

fig.2 Comparison of Model Accuracy

The mean average precision (MAP) of RetinaNet reached 82.89%, but the frames per second (FPS) is only one-third of YOLO v5, which makes it difficult to achieve real-time performance. SSD does not perform as well on the indicators of MAP and FPS. RetinaNet is known for its accuracy and the ability to handle objects at various scales. It is particularly strong in scenarios where high accuracy is critical.

In terms of accuracy, if the primary goal is to achieve the highest accuracy in object detection, RetinaNet may be the better choice. It's often used in scenarios where precision is critical.

However, YOLOv5 has made significant strides in improving accuracy while maintaining its speed, making it a strong choice for real-time applications where a good balance between speed and accuracy is required.

YOLOv5 has the accuracy of detecting objects is 90%, this is the best model for real-time object detection. YOLOv5 is known for its speed and efficiency. It can provide real-time or near-real-time object detection, making it suitable for applications that require quick processing. For smart solid waste management real time detection is required and it is must, so YOLOv5 is the best choice as it has the good accuracy and speed.

The confusion matrix of the YOLOv5 for the metal can detection is shown in fig.3 from the confusion matrix, one can calculate the false positive rate, recall, f1 score and precision. These values are false positive rate=0.4, precision=0.57, recall=0.67 and f1-score=0.62.

The authors in [8] aims to enhance ship detection in challenging scenarios, including arbitrary ship orientations, large aspect ratios, and densely arranged ships in remote sensing images. The proposed method includes: Rotated RetinaNet: This component achieves rotation detection using a feature pyramid network, rotated anchors, skew intersection-over-union (IoU), and skew non-maximum suppression.

Refined Network: An additional network improves detection accuracy. Feature Alignment Module: This module enhances feature alignment to boost detection performance.

Improved Loss Function: The loss function is enhanced by introducing an IoU constant factor to address boundary discontinuity.

The paper introduces a new dataset with more accurate labels, additional images, and diverse object samples. Through an ablation study, the effectiveness of the rotated RetinaNet, feature alignment module, and improved loss function is thoroughly analyzed. The experimental results demonstrate that the proposed method outperforms other state-of-the-art techniques in ship detection, making it a valuable contribution for remote sensing image analysis.

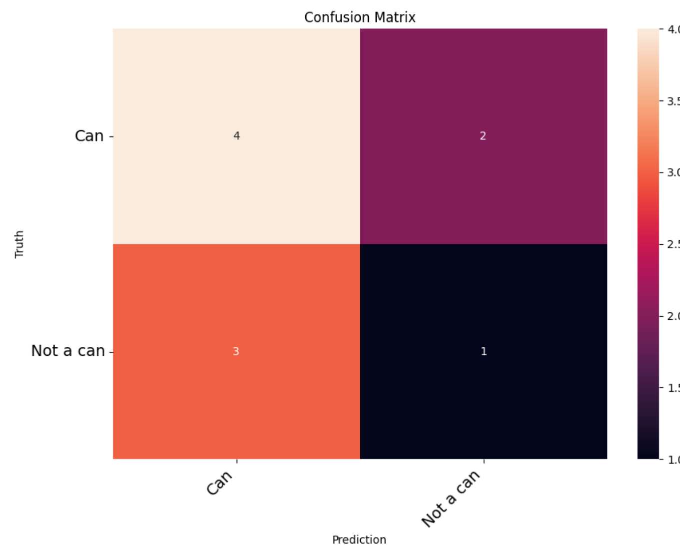

fig.3 Confusion Matrix of Yolov5

V. CONCLUSION

This research findings indicate that ML algorithms have a high potential for object detection. This study of 4 models found that RetinaNet has the highest accuracy, but the computational power and processing speed is very slow and its fps is one-third of the YOLOv3. YOLOv5 is the upgraded version of YOLOv3. It has the second-highest accuracy, high computational power and fast processing speed making it the best suitable model for real-time object detection. In terms of accuracy, if the primary goal is to achieve the highest accuracy in object detection, RetinaNet is the best choice. It's often used in scenarios where precision is critical. The choice between RetinaNet and YOLOv5 hinges on the specific requirements of the task at hand. If the primary goal revolves around achieving the utmost accuracy in object detection, RetinaNet stands out. It is commonly favoured in situations where precision and detailed object localization are critical. Conversely, for real-time object detection scenarios demanding a balance between accuracy and processing speed, YOLOv5 shines as a robust solution. Ultimately, this research underscores the importance of selecting the right object detection model based on the specific needs of the application.

REFERENCES

[1] L. Zhou et al., ”SWDet: Anchor-Based Object Detector for Solid Waste Detection in Aerial Images,” in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 306-320, 2023, doi: 10.1109/JSTARS.2022.3218958.

[2] A. Womg, M. J. Shafiee, F. Li and B. Chwyl, ”Tiny SSD: A Tiny Single-Shot Detection Deep Convolutional Neural Network for Real-Time Embedded Object Detection,” 2018 15th Conference on Computer and Robot Vision (CRV), Toronto, ON, Canada, 2018, pp. 95-101, doi: 10.1109/CRV.2018.00023.

[3] Al-Akhir Nayan, Joyeta Saha , Khan Raqib Mahmud , Abul Kalam Al Azad , Muhammad Golam Kibria.(2020).“Detection of Objects from Noisy Images”. International Conference on Sustainable Technologies for Industry.

[4] Jihun Yoon, Seungbum Hong, and Min-Kook Choi .(2021).“Semi-Supervised Object Detection With Sparsely Annotated Dataset”.IEEE

[5] Ye Tao., Zhao Zongyang ., Zhang Jun ., Chai Xinghua., Zhou Fuqiang.(2021). “Low-altitude small-sized object detection using lightweight feature-enhanced convolutional neural network”. Journal of Systems Engineering and Electronics

[6] Lubna Aziz., Md. Sah Bin Haji.,Salam.,Usman Ullah Sheikh.,Sara Ayub.(2020)”exploring Deep Learning-Based Architecture, Strategies, Applications and Current Trends in Generic Object Detection: A Comprehensive Review”. IEEE Access,10.

[7] Junhyung Kang ., Shahroz Tariq.,Han Oh 4 ,Simon S. Woo.,(2022).“A Survey of Deep Learning-Based Object Detection Methods and Datasets for Overhead Imagery” .IEEE Access,10.

[8] M. Zhu et al., ”Arbitrary-Oriented Ship Detection Based on RetinaNet for Remote Sensing Images,” in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 6694-6706, 2021, doi: 10.1109/JSTARS.2021.3082526.

[9] Sean Brown, Caitlin Hall, Raffaele Galliera and Sikha Bagui(2022).”Object Detection and Ship Classification Using YOLOv5”BOHR International Journal of Computer Science 2022, Vol. 1, No. 1, pp. 124–133.