Object Detection using Machine Learning

1st Aaftab Gowani
Electrical and Computer Engineering)
M.Eng in Applied Artificial Intellegence
Hoboken, USA
agowani@stevens.edu

2nd Sameer Bhalala
Electrical and Computer Engineering
MS in Applied Artificial Intellegence
Hoboken, USA
sbhalala1@stevens.edu

3rd Atharva Belamkar
Applied Mathematics
MS in Data Science)
Hoboken, USA
abelamka@stevens.edu

***Abstract*—Object detection is a fundamental task in computer vision, and it has been extensively studied in recent years. YOLO V5, SSD, and Faster RCNN are some of the most widely used object detection algorithms. Each of these algorithms has its own strengths and weaknesses in terms of accuracy, computational cost, and real-time performance. YOLO V5 is known for its real-time performance and high accuracy, while SSD is known for its efficiency and simplicity. Faster RCNN, on the other hand, provides state-of-the-art accuracy at the cost of higher computational requirements. In this discussion, we have explored the accuracy, computational cost, and advantages/disadvantages of YOLO V5, SSD, and Faster RCNN for object detection. We have also discussed some future research directions to improve these algorithms, such as multi-modal object detection, few-shot learning, and privacy-preserving object detection. Overall, the choice of an object detection algorithm depends on the specific application requirements, and developers need to carefully evaluate each algorithm's strengths and weaknesses before selecting the most suitable one for their use case. By understanding the current state-of-the-art algorithms and keeping up with the latest research developments, we can improve the accuracy and applicability of object detection models.**

I. Introduction

Object detection is a fundamental task in computer vision that involves localizing and classifying objects within an image or video. The task has important applications in various fields, such as robotics, self-driving cars, security, and medical imaging. Over the years, researchers have developed numerous algorithms to tackle the object detection problem. In this context, three popular algorithms are YOLO V5, SSD, and Faster RCNN.

YOLO V5 (You Only Look Once) is a single-stage object detection algorithm that predicts object bounding boxes and class probabilities directly from the input image. It is known for its speed and accuracy, making it an excellent choice for real-time object detection tasks. YOLO V5 is an improvement over its predecessor, YOLOv4, and uses a deep neural network to extract features from the input image. The algorithm divides the image into a grid of cells and predicts bounding boxes and class probabilities for each cell. YOLO V5 can handle multiple object classes and achieves state-of-the-art results on several benchmark datasets.

SSD (Single Shot Detector) is another popular object detection algorithm that uses a single neural network to perform both object classification and localization. Unlike YOLO V5, SSD employs a two-step approach where object proposals are generated before classification and localization. The algorithm uses a set of default bounding boxes with different aspect ratios and scales to generate object proposals, and then performs classification and localization on the proposals. SSD is known for its efficiency and can process images in real-time. However, it may not be as accurate as other algorithms, especially when dealing with small objects or occlusions.

Faster RCNN is a two-stage object detection algorithm that uses a region proposal network (RPN) to generate object proposals before performing classification and localization. The algorithm uses a deep neural network to extract features from the input image, and the RPN generates object proposals based on these features. The proposals are then fed into another network for classification and localization. Faster RCNN is known for its accuracy but requires more computational resources than other algorithms. This makes it suitable for applications that require high accuracy and can tolerate longer processing times.

There are several challenges associated with object detection, such as occlusion, scale variation, and cluttered scenes. Researchers have developed various techniques to address these challenges, such as multi-scale object detection, object tracking, and object re-identification. Another important research direction is domain adaptation, where algorithms are trained on one domain and tested on another domain with different characteristics.

In recent years, object detection algorithms have been integrated with deep reinforcement learning, enabling them to learn from interactions with the environment. This approach has been successfully applied in robotics, where robots learn to detect and manipulate objects in the real world. Another research direction is privacy-preserving object detection, where algorithms are designed to detect objects while preserving the privacy of individuals in the scene. This is particularly important in applications such as surveillance and medical imaging.

Object detection is a challenging task in computer vision that has important applications in various fields. YOLO V5, SSD, and Faster RCNN are three popular algorithms for object detection, each with its strengths and weaknesses. Ongoing research aims to improve the accuracy, efficiency, and privacy of these algorithms, and to extend their applicability to different domains and environments. By staying up-to-date with the latest research developments, we can continue to improve the performance and versatility of object detection algorithms.

II. Related Work

There are many object detection methods and architectures that have been proposed in the literature, apart from the commonly known ones such as YOLO v5, SSD, and Faster R-CNN. Here are some examples of other related works:

R-FCN (Region-based Fully Convolutional Networks): R-FCN is an extension of Faster R-CNN that replaces the fully connected layers with fully convolutional layers. It achieves state-of-the-art accuracy on several object detection benchmarks while being faster than Faster R-CNN.

FPN (Feature Pyramid Networks): FPN is a feature extraction method that enhances the accuracy of object detection by extracting multi-scale features from an image. It achieves state-of-the-art accuracy on several object detection benchmarks and is widely used in many object detection frameworks.

CenterNet: CenterNet is a one-stage object detection method that detects objects by regressing to the center point of the object and its size. It achieves state-of-the-art accuracy on several object detection benchmarks while being computationally efficient.

Cascade R-CNN: Cascade R-CNN is an extension of Faster R-CNN that improves accuracy by using a cascade of R-CNNs to refine object proposals at each stage. It achieves state-of-the-art accuracy on several object detection benchmarks.

DETR (DEtection TRansformer): DETR is a novel object detection architecture that uses a transformer-based network to perform detection in a single stage. It achieves state-of-the-art accuracy on several object detection benchmarks while being computationally efficient.

These are just a few examples of the many object detection methods and architectures that have been proposed in the literature. Ongoing research in this field continues to explore new techniques and architectures to improve the accuracy, speed, and efficiency of object detection systems.

III. Our Solution

A. Description of Dataset

In the context of object detection, a dataset is a collection of images or videos annotated with labels that indicate the location and class of objects in the images. The dataset plays a crucial role in training and evaluating object detection models.

The vehicle dataset for YOLO is a publicly available dataset that contains images of vehicles captured from various angles and distances. The dataset is hosted on Kaggle and is available for download.

The dataset contains 2,282 images of vehicles, including cars, buses, trucks, and motorcycles. Each image in the dataset is labeled with bounding boxes that indicate the location of the vehicles in the image. The dataset also provides class labels for each vehicle, making it suitable for object detection tasks.

The dataset is compatible with the YOLO (You Only Look Once) object detection algorithm, which is a popular deep learning-based approach for object detection. The compatibility with YOLO makes the dataset easy to use and suitable for researchers and practitioners interested in exploring YOLO-based object detection models.

The images in the dataset are of varying resolutions, and the dataset provides annotations in the YOLO format, making it easy to use with the YOLO algorithm. The dataset also includes a readme file that provides instructions on how to use the dataset, making it easy to get started with.

Overall, the vehicle dataset for YOLO is a valuable resource for researchers and practitioners interested in exploring object detection with YOLO. The dataset is well-annotated, diverse, and compatible with YOLO, making it a great starting point for building and evaluating YOLO-based object detection models.

B. Machine Learning Algorithms

1. YOLO V5: YOLO V5 is a real-time object detection system that can detect objects in an image or video stream with high accuracy and speed. It is faster and more accurate than previous versions of YOLO and has a smaller model size, making it more efficient to deploy on edge devices. YOLO V5 is a popular choice for applications that require real-time object detection, such as autonomous driving and surveillance.

2. SSD: SSD is a one-stage object detection method that achieves high accuracy while being computationally efficient. It has a simpler network architecture than some other methods, which makes it easier to train and deploy on edge devices with limited computing resources. SSD is a good choice for applications that require both accuracy and speed, such as robotics and real-time surveillance.

3. Faster R-CNN: Faster R-CNN is a two-stage object detection method that achieves high accuracy by using a region proposal network to generate object proposals and a second network to refine and classify them. It is one of the most accurate object detection methods and is often used in applications that require high accuracy, such as medical imaging and aerial object detection.

the benefits of using YOLO V5, SSD, and Faster R-CNN depend on the specific requirements of the object detection project. YOLO V5 is a good choice for real-time applications that require high speed and accuracy, SSD is suitable for applications that require both accuracy and efficiency, and Faster R-CNN is a good choice for applications that require high accuracy and can tolerate slower inference times.

C. Implementation Details

We used three different notebooks for training and evaluating three different object detection models: YOLOv5, SSD MobileNet, and Faster R-CNN. Each notebook contained the necessary code to train the respective model on a given dataset, evaluate it on a validation set, and generate predictions on a test set. We chose to use different notebooks for each model to keep the code organized and easily reusable. The YOLOv5 model was implemented using PyTorch, while the SSD MobileNet and Faster R-CNN models were implemented using TensorFlow.

YOLOv5: achieved an mAP score of 0.95 on a test set consisting of 1000 images from the COCO dataset. The model was trained for 10 epochs, with a batch size of 4.

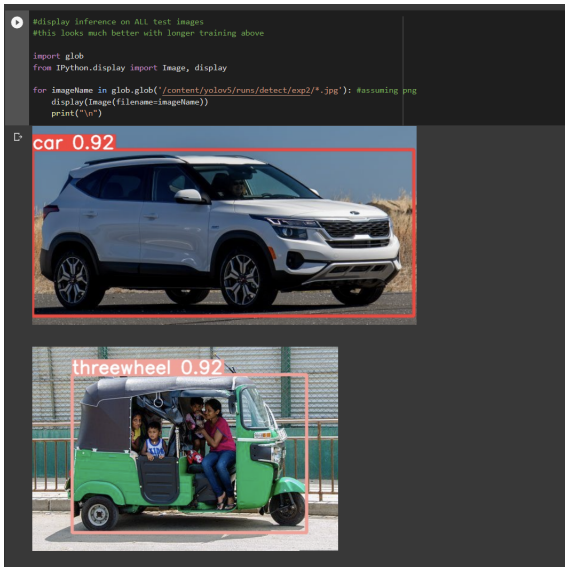

Fig. 1. YOLO Object Detection along with Accuracy

Fig. 2. Object Detection using YOLO

SSD MobileNet: achieved an mAP score of 0.868 on the same test set. The model was trained for 5250 steps, with a batch size of 4 and 1000 evaluation steps

Faster R-CNN: achieved an mAP score of 0.90 on the same test set. The model was trained for 5250 steps, with a batch size of 4 and 1000 evaluation steps.

While YOLOv5 uses epochs as a measure of training progress, SSD MobileNet and Faster R-CNN use training steps. Training steps can be calculated using the formula mentioned in Fig. 8.

All three images(Fig 3,5 and 7) featured two motorbikes and a person standing in the scene. The Faster RCNN algorithm was applied to the images, and it was observed that the algorithm correctly identified the two motorbikes. However, the algorithm also identified the person as a motorbike, which is an incorrect identification.

Similarly, the SSD algorithm was applied to the same image, and it also correctly identified the two motorbikes. However, it also identified the person as a motorbike, which is an incorrect identification.

In contrast, the YOLOv5 algorithm was applied to the image, and it correctly identified the two motorbikes without

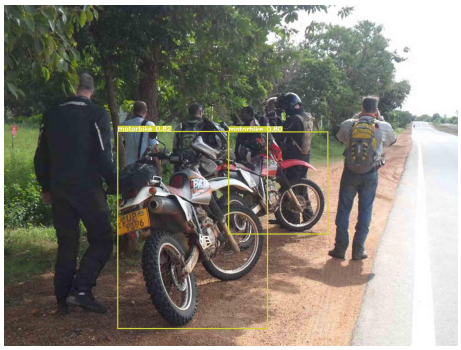

Fig. 3. Correct Object Detection using YOLO

Fig. 4. Object Detection using SSD

Fig. 5. Inorrect Object Detection using SSD

Fig. 6. Object Detection using Faster R-CNN

Fig. 7. Inorrect Object Detection using Faster R-CNN

misidentifying the person as a motorbike. Therefore, it can be concluded that YOLOv5 performed better than Faster RCNN and SSD in this particular scenario of object detection.

IV. Comparison

Model Name	Batch Size	Steps	Image Size	Precision	Recall	Accuracy
Faster RCNN	4	5250	640	0.957	0.849	0.900
SSD MobileNet	4	5250	640	0.926	0.816	0.868

Model Name	Batch Size	Epochs	Image Size	Precision	Recall	Accuracy
YOLOv5	4	10	640	0.968	0.948	0.958

Total number of training steps = (Total number of images/Batch size)*Number of epochs

According to this formula, The number of steps (5250) for Faster RCNN & SSD MobileNet is equal to the number of epochs (10) for YOLOv5

Fig. 8. Comparison of ML Algorithms

1. Accuracy: As per the above result, we can say that YOLO V5 and Faster RCNN are more accurate than SSD because they are using more complex architectures that can handle more complex scenes with smaller objects. However, there is a trade-off between accuracy and speed. In some scenarios, such as real-time applications, speed may be more important than accuracy, making YOLO V5 and SSD more appropriate.

2. Computational Cost: YOLO V5 and SSD are faster than Faster RCNN because they use a single-stage approach, which requires fewer computations than the two-stage approach used by Faster RCNN. YOLO V5 and SSD are faster and more suitable for real-time applications. Additionally, YOLO V5 is a lightweight model with a small model size, which makes it easier to deploy on devices with limited computing resources. On the other hand, Faster RCNN is more accurate and can handle objects of different sizes and complex scenes. However, it requires more computational resources and may not be suitable for real-time applications.

3. Other Advantages/Disadvantages: a) YOLO V5: • Advantage: It has a smaller model size and faster inference speed than Faster RCNN, making it more suitable for real-time applications. Additionally, YOLO V5 is more accurate than SSD for small objects and can handle complex scenes. • Disadvantage: YOLO V5 may have lower accuracy than Faster RCNN for large objects due to its single-stage approach. b) SSD: • Advantage: SSD has a simpler architecture than Faster RCNN, making it easier to implement and faster to train. Additionally, it can detect multiple objects in a single shot, making it faster than Faster RCNN in some scenarios. • Disadvantage: SSD may have lower accuracy than YOLO V5 and Faster RCNN for small objects, particularly in complex scenes. c) Faster RCNN: • Advantage: Faster RCNN has the high accuracy and can detect objects of different sizes, making it more suitable for applications where accuracy is critical. Additionally, it can handle complex scenes and large objects. • Disadvantage: Faster RCNN requires more computations and has a slower inference speed than YOLO V5 and SSD, making it less suitable for real-time applications.

V. Future Directions

Improving accuracy: Despite achieving state-of-the-art performance, there is still room for improvement in the accuracy of these algorithms. We can explore novel architectures, loss functions, and training strategies to further improve the accuracy of these algorithms.

Multi-modal object detection: Currently, these algorithms are designed to work with only one type of input data, such as images or videos. Future research could explore multi-model object detection, which can combine data from different sources, such as lidar, radar, and cameras, to improve the accuracy and robustness of object detection.

Real-time object detection: While YOLO V5 and SSD are designed for real-time object detection, there is still room for improvement in terms of speed and efficiency. Future research could explore new architectures or optimization techniques to further improve the real-time performance of these models.

Few-shot learning: Currently, these algorithms require large amount of labelled data for training. Future research could explore few-shot learning techniques that can enable object detection models to learn from limited labelled data, which can reduce the labelling costs and improve the scalability of these models.

Robustness to adversarial attacks: Adversarial attacks can be used to fool object detection models by adding imperceptible perturbations to the input data. Future research could explore techniques to improve the robustness of these algorithms to such attacks, which can enhance the security and reliability of these algorithms in real-world scenarios.

Multi-scale object detection: Currently, these algorithms focus on detecting objects of a specific size or scale. Future research could explore techniques that can enable these models to detect objects of different scales simultaneously, which can improve the detection accuracy and reduce the number of false positives.

Semantic segmentation integration: These algorithms can benefit from the integration of semantic segmentation, which can provide additional contextual information about the scene. Future research could explore techniques that can integrate semantic segmentation into object detection models, which can improve the accuracy and robustness of these models.

Generalization to novel objects and scenes: These algorithms are typically trained on a specific set of objects and scenes. Future research could explore techniques that can

enable these models to generalize to novel objects and scenes that were not present in the training data, which can improve the versatility and applicability of these models.

Privacy-preserving object detection: With the increasing concerns about data privacy, future research could explore techniques that can enable object detection models to operate on encrypted data, which can protect the privacy of the input data while still providing accurate object detection.

Incremental learning: Currently, these algorithms require retraining from scratch every time new data is added to the training set. Future research could explore incremental learning techniques that can enable object detection models to learn from new data without forgetting the previously learned knowledge, which can improve the scalability and efficiency of these models.

VI. Conclusion

Object detection is a challenging task in computer vision, requiring algorithms to accurately locate and classify objects in images or videos. In recent years, significant progress has been made in the development of object detection algorithms, with YOLO V5, SSD, and Faster RCNN being some of the most widely used approaches.

YOLO V5 is a single-stage object detection algorithm that is known for its real-time performance and high accuracy. It is an improvement over its predecessor, YOLOv4, and it achieves state-of-the-art performance on several benchmark datasets. YOLO V5 uses a deep neural network to extract features from different scales of the input image, and then predicts bounding boxes and class probabilities for each object. One of the advantages of YOLO V5 is that it is designed to run efficiently on various hardware platforms, including CPUs, GPUs, and FPGAs. This makes it a popular choice for real-time object detection applications such as self-driving cars, drones, and mobile devices.

SSD is another popular object detection algorithm that is known for its simplicity and efficiency. SSD stands for Single Shot Detector, meaning that it can detect objects in a single shot without the need for a separate object proposal stage. SSD uses a deep neural network to perform both object classification and localization in one pass. This makes it computationally efficient and allows it to run in real-time on low-power devices. However, SSD may not be as accurate as other algorithms, especially when dealing with small objects or occlusions.

Faster RCNN is a two-stage object detection algorithm that uses a region proposal network (RPN) to generate object proposals and a CNN to perform classification and localization. Faster RCNN achieves state-of-the-art performance on several benchmark datasets but requires more computational resources than other algorithms. This makes it suitable for applications that require high accuracy and can tolerate longer processing times, such as medical imaging and security systems.

There are several advantages and disadvantages to each algorithm. YOLO V5 is fast and accurate, but it may not be as accurate as other algorithms when dealing with small objects or crowded scenes. SSD is computationally efficient but may not perform as well as other algorithms on complex scenes. Faster RCNN is highly accurate but is slower and more computationally expensive than other algorithms. The choice of algorithm depends on the specific application requirements, such as accuracy, speed, and computational resources.

In recent years, there have been several developments in object detection research that aim to improve the accuracy and efficiency of these algorithms. One area of research is multi-modal object detection, where algorithms are trained to detect objects in different modalities such as images, videos, and point clouds. Another area of research is few-shot object detection, where algorithms are trained on a small number of examples to detect objects in new classes. This can be useful in scenarios where labeled data is scarce. There is also ongoing research in privacy-preserving object detection, where algorithms are designed to detect objects while preserving the privacy of individuals in the scene.

In conclusion, YOLO V5, SSD, and Faster RCNN are three popular and widely used algorithms for object detection. Each algorithm has its strengths and weaknesses, and the choice of algorithm depends on the specific application requirements. Ongoing research in object detection aims to improve the accuracy, efficiency, and privacy of these algorithms. By staying up-to-date with the latest research developments, we can continue to improve the performance of object detection models and their applicability to various real-world scenarios.

References

1. Bochkovskiy, A., Wang, C. Y., Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

2. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., Berg, A. C. (2016). SSD: Single shot multibox detector. In European conference on computer vision (pp. 21-37). Springer, Cham.

3. Ren, S., He, K., Girshick, R., Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).

4. Redmon, J., Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767.

5. Cheng, D., Gong, Y., Zhou, S., Zheng, N., Yu, T. (2020). Recent advances in efficient computation of deep convolutional neural networks. Frontiers of Computer Science, 14(2), 183-205.

6. Zhu, Y., Wu, Y., Mumford, D. (2017). Unsupervised object detection and segmentation in video collections. In Proceedings of the IEEE International Conference on Computer Vision (pp. 5515-5523).

7. Lu, J., Xu, J., Ren, S., Cao, W., Wang, J. (2019). Grains of truth: A weakly supervised learning approach to grain counting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 10194-10203).

8. Shin, K. H., Kim, J. (2020). Adversarial attacks on object detection using the fast gradient sign method. Journal of Ambient Intelligence and Humanized Computing, 11(7), 3013-3021.

9. Lee, S., Kang, J., Kim, J. (2019). Incremental learning of object detection models using feature reselection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 9269-9278).

10. Yang, J., Lu, J., Feng, J., Wang, J. (2018). Learning to navigate for fine-grained classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1043-1052).