www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 02, February 2025, pp : 1315-1322

e-ISSN : 2583-1062
Impact Factor : 7.001

A SURVEY OF DEEP LEARNING ARCHITECTURES FOR OBJECT DETECTION IN COMPUTER VISION

Dr. Neha Yadav[*1], Chaitanya Sharma[*2], Gauri Gandhi[*3], Divy Pant[*4]

[*1]Assistant Professor, Department Of Artificial Intelligence And Machine Learning, ADGIPS, Delhi, India.

[*2,3,4]Scholar, Department Of Artificial Intelligence And Machine Learning, ADGIPS, Delhi, India.

DOI: https://www.doi.org/10.58257/IJPREMS38732

ABSTRACT

Object Detection is a prominent area in computer vision, where deep learning has dramatically advanced in many areas-from autonomous driving and healthcare to surveillance. Discuss the development of deep learning models for object detection: two-stage detectors like Faster R-CNN, one-stage detectors as YOLO and SSD, and emerging transformer-based models like DETR. We discuss strengths and weaknesses of each type of model with respect to accuracy, speed, and efficiency of resources used, specifically looking at the challenges such models pose in real applications like occlusion, detection of small objects, and domain adaptation. Finally, we describe how large datasets like MS COCO and PASCAL VOC became important to the development of benchmarks. Future promising research directions would be multi-modal learning, lightweight models for resource-constrained devices, and ethics considerations for privacy-sensitive applications. This review tries to outline the state-of-the-art object detection methodology available nowadays, indicates the challenges of the present situation, and points out how further development might occur.

Keywords: Computer Vision, Deep Learning, R-CNN, YOLO, SSD, DETR, MS COCO, PASCAL VOC, Multi-Modal Learning.

1. INTRODUCTION

The past few years have seen great revolutions in computer vision with the development of deep learning. This has opened immense spaces for image classification, segmentation, and object detection. Among the problems that define the computer vision challenge is the recognition and localization of objects in images. Often considered the most important part of this subset, object detection has been in the spotlight lately due to its potential applications along with model developments achieved because of advances in deep learning.

These CNNs have revolutionized the field to a large extent. They had introduced learning paradigms end-to-end without having to hand-engineer features directly from raw pixels, thus optimizing the process of object detection. The more advanced architectures of the Faster R-CNN, YOLO, and SSD render great trade-offs between detection speed and accuracy in an object detection.

Furthermore, the existence of big, annotated datasets like MS COCO, PASCAL VOC, and ImageNet has acted as a catalyst in promoting development. These provide standardized benchmarks to support the evaluation and comparison of different object detection models and stimulate innovation.

This survey paper is an overview of the main architectures of deep learning that have been used for object detection. Its focus is on pointing out contributions, applications, and challenges that remain open today.

2. BACKGROUND AND KEY CONCEPTS

Object detection is an important aspect of visual recognition in computer vision that involves identifying and localizing instances of objects within an image through bounding boxes. Contrast this to object classification, where it simply gives the category of the object, and object detection is much more complex and computationally expensive as it involves very good spatial localization. Applications of object detection include but are not limited to: autonomous vehicles, healthcare, surveillance, and robotics.

Deep representation with the application of Convolutional Neural Networks (CNNs) has radically advanced object detection. CNN is built to infer spatial hierarchies in visual data by learning different abstraction layers, and this has been highly effective for object detection at various scales and orientations. In a typical CNN architecture, convolutional, pooling, and fully connected layers allow those stages to contribute to the network's ability to learn and generalize features. Convolutional layers detect visual features, such as edges and textures. Pooling reduces the dimensionality of such features without losing important information. Fully connected layers enable classification or regression with respect to the detected features. Two primary categories of object detectors in deep learning approaches are one-stage detectors and two-stage detectors.

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 02, February 2025, pp : 1315-1322

e-ISSN : 2583-1062
Impact Factor : 7.001

Two-Phase Detectors: These models include Faster R-CNN and Mask R-CNN. Object detection is performed in two steps. In the first step, it produces regional proposal images that presumably contain objects. The second phase refines the proposals by object classification as well as adjustment of bounding boxes. The model gives good accuracy but incurs significant computational complexity. Thus, this model is preferable over the applications where speed is less Important.

Other examples of one-stage detectors are those that predict at once, in a single step, both bounding boxes and object categories, such as YOLO and SSD. They appear to be much faster than the above two-stage detectors and are especially well suited to real-time requirements. By contrast, one-stage detectors tend to be slightly less accurate than their competitors for the tasks of detecting smaller objects.

But the second part of object detection, which evaluates both the accuracy and localization of the model, is the evaluation metrics. Among such, a few common ones in usage are Precision, Recall, F1-Score, and mean Average Precision. Out of these, the highly useful metric is mAP, since it calculates average precision across all objects in every category, making it easy to comprehensively compare models.

Understanding these building blocks—CNN architecture, object detection frameworks, and evaluation metrics—lays the groundwork for delving into more advanced models and nascent trends in the field.

3. DEEP LEARNING MODELS FOR OBJECT DETECTION

Fluid Deep learning architecture was significantly improved to maximize the performance of object detection. Such improvement produced different architectures tailored for either accuracy, speed, and computational efficiency. There are basically two types of deep learning-based object detection models: two-stage detectors and one-stage detectors.

a. Two-Stage Detectors

Two-stage detectors include two steps; the first stage generates a set of regional proposals possibly containing objects which are further classified and refined to precisely locate and categorize each object in the second stage. This is essentially a good approach toward high detection accuracy where careful localization of objects is required in complex scenes.

- R-CNN and its Variants: The first two-stage detector was the R-CNN (Regions with CNN features). These utilized selective search to generate the regions, which then were passed to the CNN for classification. Variants of Fast R-CNN shared convolutional features across regions to save computation but eliminated the necessity of using separate region proposal algorithms by introducing Region Proposal Network in Faster R-CNN, making the whole process end-to-end trainable.
- Mask R-CNN: Mask R-CNN is an extension of the Faster R-CNN by adding a segmentation branch that predicts object masks, apart from bounding boxes and labels. This innovation allows Mask R-CNN to carry out instance segmentation, which makes it very useful in the presence of applications requiring detailed information about the shapes of the detected objects. Two-stage detectors are very accurate but generally slower as the computation of detection is sequential. Hence, this two-stage detector will be relatively good at applications that primarily need a high accuracy of detection and do not shun computational abilities, such as medical imaging or advanced robotics.

3.2 Single-Stage Detectors

Single-stage detectors are intended for real-time applications along the mainline of simplifying the detection process into a single step that directly predicts object bounding boxes and class probabilities over the entire image in a single forward pass. At the cost of losing perhaps a little bit in terms of accuracy, they achieve speeds significantly higher than two-stage detectors and thus are highly desirable in applications where real-time performance matters.

- YOLO (You Only Look Once): YOLO transformed object detection into just one problem of regression. The network split an image into a grid, and for each cell, the model predicted bounding boxes along with class probabilities, hence increasing the speed of detection to orders of magnitude. Subsequent versions, including YOLOv3 and YOLOv4, achieved more accuracy but retained efficiency and proved suitable for applications like surveillance, autonomous driving, etc.
- SSD (Single Shot MultiBox Detector): SSD added multi-scale feature maps; therefore, it could now detect objects at any scale with precision. It, like YOLO, does single pass-through images but strikes a balance between preciseness and efficiency in detecting smaller objects more precisely. The simple and efficient design of SSD makes it widely used in mobile and embedded devices.

3.3 Emerging Architectures and Innovations

The recent development of deep learning generated interest in new architectures and hybrid models where the effectiveness of object detection can be boosted: Transformer-based models. Motivated by the success that a

transformer achieved in NLP applications, a new model was presented, called DETR, (Detection Transformer), that relies on self-attention mechanisms for modeling long-range dependencies. This confers freedom on various forms of spatial relationships; hence this kind of model has an advantage in complex detection tasks.

- Hybrid Approaches: Other newer approaches combine the strength of CNNs and transformers, or home in on Recurrent Neural Networks (RNNs) and attention mechanisms to look after the temporal nature of the problem, like video object detection.

3.4 Conclusion

Advancements in object detection models based on deep learning have led to the availability of several variants depending on the accuracy vs. speed requirements. Models range from two-stage detectors, emphasized in terms of high precision, to one-stage detectors optimized for real-time applications. These models are probably the best examples of the versatility of deep learning in dealing with different object detection needs. This chapter gives an overview of the main architectures that form a basis for discussing specific models and their applications.

4. OBJECT DETECTION DATASETS AND BENCHMARKS

Datasets acted as a leap forward in object detection by allowing a structured way to train, test, and evaluate. It is then large, annotated datasets that would allow such models to generalize well across diverse scenes and object categories. Several benchmark datasets have played an instrumental role in driving progress in object detection.

a. Popular Object Detection Datasets

- PASCAL VOC: One of the first datasets in any working application of object detection was PASCAL Visual Object Classes (VOC). It consists of many objects in scenes of daily life and allows classification, detection, as well as segmentation. Probably the most often used versions are PASCAL VOC 2007 and 2012; thousands of images are annotated using bounding boxes and object categories.
- MS COCO: this is one of the most used datasets, purely because of its broad annotation and richness of types. This dataset includes pictures concerning more than 200,000 images provided with labels in the form of bounding boxes and instance segmentation masks. For several categories, even key points are available. The 80 object categories and complex scenes with multiple objects quickly explain why MS COCO has become a standard benchmark to work with models that develop object detection or segmentation.
- ImageNet: Although ImageNet was primarily designed for image classification, it also released object detection in the form of large numbers of images across various categories. The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) includes a detection task that pushes the limits of detection of objects in thousands of categories.
- Open Images Open Images developed by Google comprises millions of annotated images with the use of bounding boxes for 600 object classes. It includes object relationships, segmentations masks, and object hierarchies that allow it to be incredibly useful for complex tasks such as relationship detection and multi-label classification.
- DOTA (Dataset for Object Detection in Aerial Images): It is highly specialized for aerial and satellite images. DOTA has images captured by drones, satellites, and other related equipment. The dataset contains annotations for all the object classes found in aerial views like buildings, vehicles, and ships. This deals with the challenges of aerial images.

b. Features and Challenge of Datasets

Each dataset has features that impact the performance of models and their usability for certain tasks:

- Object Diversity: COCO and Open Images have humongous object classes and numerous annotations. This would push the models to learn more general representations, because of which one can use these representations on a wide variety of tasks. Datasets such as DOTA, however, are designed to be focused on specific object classes that are of special interest in a particular domain say aerial imagery.
- Image Complexity: Scenes in any dataset, such as COCO with various objects in different contexts, are worth the task of testing models' real-world object detection capabilities. Similarly, the complicated annotations within Open Images are useful for training on subtle relationships between the objects.
- Scale and Data: Sure enough, massive datasets like ImageNet and Open Images can be used to train models to generalize well. At the same time, it demands more computations during training.

c. Evaluation Metrics

Evaluation metrics standardized on different data sets enable easy comparison of different models. The basic evaluation metrics used for object detection are:

- Precision and Recall: These are metrics to gauge how accurate a model is in the identification of objects on the image (precision) and its ability to detect all relevant objects (recall). It gives a balanced view of model performance in both correctness and completeness.
- Intersection over Union (IoU): IoU measures the overlap between bounding boxes predicted and ground truth. High IoU means good localization accuracy. Typically, IoU thresholds are 0.5 or higher, which defines whether a detection is good.
- Mean Average Precision (mAP): It is probably the most widely used metric and captures the precision-recall curve for all classes in a dataset. Calculated at different IoU thresholds, aggregate mAP measure for model performance makes it the benchmark standard for object detection.

d. Summary

These datasets and metrics have driven much of the research for object detection by providing test beds on which models are developed, validated, and compared. They overcome many difficulties related to diversity in objects, complexity in scenes, and scale. The work is continuously being done to develop robust and versatile object detection models by continually advancing the stride to create such models. In this chapter, we have presented critical datasets and benchmarks that form the root of comparison for object detection models.

5. PERFORMANCE EVALUATION AND COMPARISON

The performance of object detection models is usually evaluated based on the considerations of a combined metric: accuracy, speed, and resource efficiency. Comparing these factors helps determine a suitable model for specific applications-whether high-speed applications demand real-time systems or whether precision-demanding tasks require controlled environments.

a. **Key Performance Matrics**

- Mean Average Precision (mAP): The mAP measure checks the average precision across several object classes along with a range of intersection-over-union (IoU) thresholds. High values of mAP correspond to good results; this is why mAP is the most appropriate standard measure in object detection benchmarks. Several IoUs, such as 0.5 and 0.75, are used to compute the mAP and check how a model performs in terms of localization under different conditions.
- PR Curve: Precision-Recall Curve This is a plot of the trade-off between recall (coverage of all relevant instances) and precision, that is, correct positive predictions. The curve will allow you to get a view of the performance of a model at different levels of confidence when weighing false positives against false negatives.
- Frame Per Second (FPS): For those applications in real-time detection, like autonomous vehicles or surveillance, FPS is a crucial metric. The higher FPS will make it possible to infer faster and pass on the processed model more frames per second. The fast models may compromise on accuracy, so FPS becomes an important consideration based on which speed versus precision trade-offs happen.
- IoU: The intersection over union between the predicted bounding box and the ground truth bounding box represents the area of overlap of the predicted and ground truth bounding boxes divided by the area of their union. IoU thresholds, that is usually equals to 0.5, determine whether to classify the detection as correct or not. A model with a greater IoU score exemplifies better localization accuracy, which can be critical in apps where exact positioning is crucial.

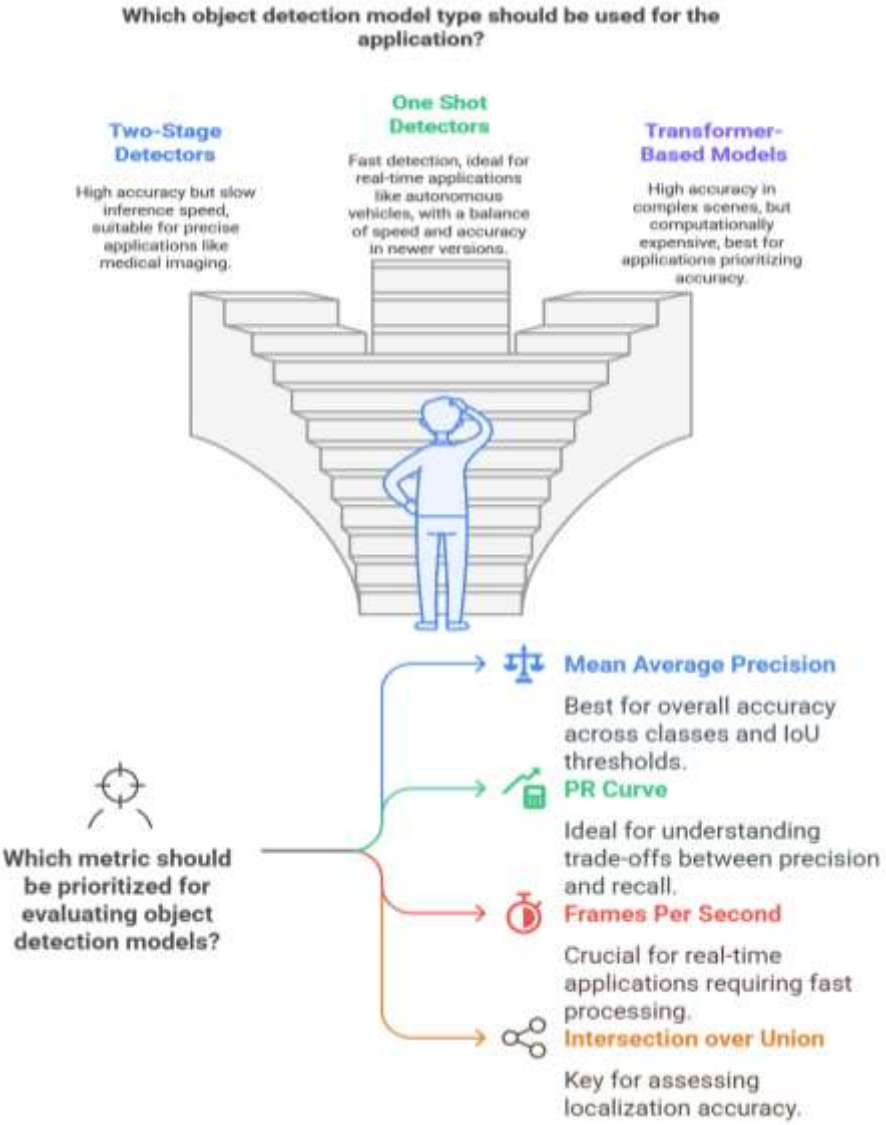
Model Comparisons

- Two-Stage Detectors: Models such as Faster R-CNN and Mask R-CNN have been found with high accuracy since they follow a two-stage method. Those models are useful where applications need greater precision to be involved in the detection process, such as medical images and quality inspection on the manufacturing side. Inference speed is low so cannot be used in real-time applications.
- One Shot Detectors: YOLO primarily deals with the variety of one-shot versions of YOLO, such as YOLOv3 and YOLOv4. They are excellent for very fast object detection in one shot. Though sometimes, they might lose some accuracy, they are highly useful for real-time applications like autonomous vehicles and surveillance. Moreover, more recent versions of models like YOLOv5 went further on the balance between speed and precision.
- Transformer-Based Models: Newest models-including DETR (Detection Transformer)-use mechanisms that allow self-attention, offer highly flexible spatial relationships, and notably improve accuracy, particularly in very complex scenes with several objects. However, computations can be expensive; therefore, these models are better suited for applications where accuracy is primary.

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 02, February 2025, pp : 1315-1322

e-ISSN : 2583-1062
Impact Factor : 7.001



b. Resource Considerations

Deploying object detection models on constrained devices, such as mobile phones and embedded systems, requires a need for efficiency in memory and computation. The most commonly used techniques to decrease model size while minimizing the inference time are model pruning, quantization, and knowledge distillation without impacting the essentially achieved accuracy level.

- Pruning: By deleting the smaller weights that are less important, it reduces the number of parameters in a model, which consequently lowers the memory usage.
- Quantization decreases the precision of calculations (for example from 32 bits to 8 bits) and faster in inference time, and also decreased the model size.
- Knowledge Distillation trains a smaller model (student model) to mimic the outputs of a more complex, high-performing model (teacher model), retaining accuracy but low computational requirements.

c. Conclusion

Choosing the best object detection model will be the specific need of the application that requires efficiency in accuracy, speed, and the constraint of resources. This comparison between two-stage and one-stage detectors, together with emergent transformer-based models, brings out the trade-offs. Performance metrics such as mAP, IoU, and FPS combined with resource efficiency techniques give a comprehensive basis for judging model suitability for different contexts. This chapter has described key performance criteria guiding model selection for object detection applications.

6. APPLICATIONS OF DEEP LEARNING IN OBJECT DETECTION

Deep learning has enabled object detection to be applied to such vast fields. Be it autonomous vehicles, medical imaging, or whatever be the application, it is the object detection models that play a very important role in making the

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 02, February 2025, pp : 1315-1322

e-ISSN : 2583-1062
Impact Factor : 7.001

automation process safer and more precise. Here are some of the most important applications of deep learning-based object detection.

a. Autonomous Vehicles

Object detection is the core part of autonomous driving systems. Real-time pedestrian, vehicle, traffic sign, and obstacles detection enable safe navigation of autonomous vehicles through dynamic environments. YOLO and SSD are among many such models, which are commonly used in the domain because they process frames at a fast rate, which is especially necessary for real-time decisions. In some cases, detectors such as Faster R-CNN are also applied wherever the requirement is high accuracy for localization but are often applied together with faster versions to balance between speed and accuracy.

b. Healthcare and Medical Imaging

Object detection in medical imaging detects tumors and fractures, as well as other pathologies, in X-rays, MRIs, and CT scans. In the case of two-stage detectors such as Faster R-CNN and Mask R-CNN, it is valuable to obtain high accuracy localization for their potential use in medical applications. Object detection has played a central role in furthering the technology for early diagnosis, surgical planning, and monitoring of treatments, thus making it highly influential in the healthcare industry.

c. Surveillance and Security

Object detection is vastly used in surveillance to detect and track people, find suspicious activities, and inform authorities in real-time. Deep learning-based object detection and facial recognition give security systems another integration and enhanced performance where identification and tracking happen even in the most complex and crowded environment.

d. Retail and Inventory Management

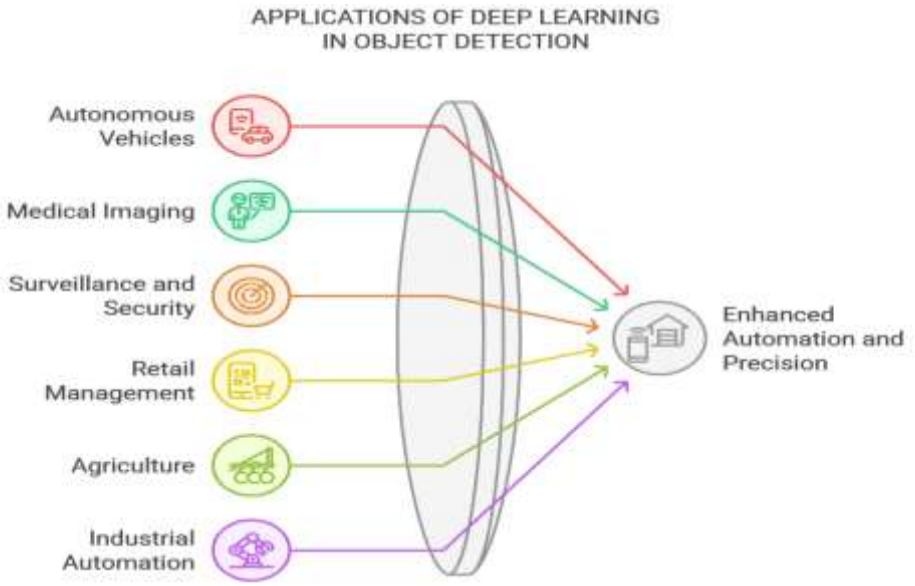
Object detection is applied in retail inventory management, checkout automation, and analyzing the behavior of customers. For instance, the automated checkout system by stores uses object detection, which does not require barcodes to check out products, among others. Beyond these, inventory management systems make use of object detection to monitor stock levels in real-time. This will allow for optimization of the supply chain and labor.

e. Agriculture

In agriculture, object detection can be used for tracking crop health, pest detection, and care for the livestock. A drone mounted with object detection models can scan large farms, search for plant diseases, troubles in soil quality, and poor-yielding crops. The high throughput of one-stage detectors such as YOLO provides a fast way of processing aerial images, which invariably is the case in the industrial agriculture sector where considerable insights need to be obtained in real-time.

f. Industrial Automation

Object detection contributes to industrial automation through quality control in the form of defect detection and sorting in production. Quality control enhances product quality, detects defectives, and reduces waste using object detection models. In addition to that, systems powered by object detection models can automatically classify items, speed up packaging processes, and monitor the functioning of equipment to inform when to perform maintenance.



www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 02, February 2025, pp : 1315-1322

e-ISSN : 2583-1062
Impact Factor : 7.001

7. CHALLENGES AND FUTURE DIRECTIONS

Deep learning-based object detection has made great progress, but several challenges remain. These challenges are meant to be overcome for better model robustness, efficiency, and applicability to diverse domains. The section below outlines major challenges in object detection and points out promising future research and development directions.

a. Challenges In Object Detection

- Handling Occlusions and Complex Backgrounds: In the real world, frequently objects are occluded or merge with complex backgrounds, which may cause some difficulties for detection. Detection of partially occluded objects requires good generalization capabilities and keeping context-of course challenging work, especially for those one-stage detectors that focus on speed.
- Detection of small objects: The greatest difficulty is detection of small objects because in a crowded scene, there is not much resolution for small objects available. Small objects occupy fewer pixels that models find hard to detect. This challenge has been partially overcome by means of multi-scale detection algorithms by feature pyramids, and hence further improvement is required to detect more precisely small objects.
- Real-Time Processing Constraints There are several applications including autonomous driving and surveillance video that require real-time object detection models. For these types of applications, one-stage detectors like YOLO or SSD certainly entail a loss in accuracy, especially for complex scenes, but still offer to achieve this speed. It remains challenging to find high accuracy with real-time performance.
- Domain Adaptation: The object detection models learn specific datasets but do not extend well to new environments, including varying scales, lighting conditions, or object appearances in the image. There's a huge gap here-a model is needed that adapts itself to new domains with minimal re-training or data collection.
- Data Privacy and Ethical Issues Though object detection is increasingly becoming important, ethical concerns regarding data privacy surveillance appear. Applications to facial recognition and surveillance may need to be approached with privacy considerations against misuse, a state intended to protect individual rights.

b. Future Directions

- Use of Transformers: Object detection can also be seen in the use of transformer-based models, for example DETR that depend upon self-attention mechanisms to capture complex spatial relationships. Hybrid models between transformers and CNNs might be an area of future work in designing both higher localization accuracy as well as computational efficiency.
- Lightweight Model Optimization: Lightweight models are required for both mobile applications and applications on embedded systems. Extending the pruning, quantization, and distillation techniques employed so far, further reduction of model size and computations would make object detection accessible on devices having very limited resources.
- Multi-Modal Object Detection: Adding depth information to images, infrared images, and so on can improve the detection of objects, particularly in difficult scenarios or environments with little lighting or occlusion. In fact, multi-modal models are the combination of visual data with other sensory sources for high-preciseness and robustness.
- Self-Supervised and Few-Shot Learning: Collecting large datasets annotated is expensive and time-consuming. Self-supervised and few-shot learning aim at training object detection models using as few annotations as possible to reduce the dependency on large datasets pre-annotated. This may further improve performance when data are scarce, helping to support faster model deployment.
- Ethically Responsible and Transparent AI: It is the time that the models of object detection must be applied ethically because models are going to be used in applications with sensitive fields, for example, surveillance. The work going on creating explainable and transparent models will ensure that the users at the end know what the model decided. Thus, it builds trust and strengthens the sense of accountability in AI-dependent systems.

8. CONCLUSION

The last few years have been excellent for the field of deep learning-based object detection. Improvements in model architecture, datasets, and computation power have pushed the state-of-the-art into significantly new directions. Techniques have appeared capable of not only rivaling accuracy but also rivaling speed, such as Faster R-CNN, YOLO, or SSD, opening broad applications across industries. Despite these achievements, there are still some challenges that need to be addressed. Some of the issues with the current state of object detection include further improving on dealing with occlusions and keeping optimal performance with increased computation efficiency in real-time applications. These deficiencies highlight the need for continuous innovation that would broaden the applicability and effectiveness of object detection models.

Future research directions would include, but not be limited to, transformer-based models, lightweight architecture for mobile deployment, and multi-modal detection techniques. All these promise better avenues in improving the performance of the model. For example, transformer architectures have already been shown to improve the spatial understanding of a model, and progress on self-supervised and few-shot learning approaches reduces the dependency on large, annotated datasets. In the end, all these bets can be looked upon as opening object detection to wider scenarios in terms of settings, availability, adaptability, and efficiency.

Ethics is yet another important aspect of future work. With the increased deployment of object detection in sensitive applications such as surveillance and healthcare, this is important so that models are clear, responsible, and aligned with privacy norms. As the research community and practitioners push for more interpretable models, the field stands to stand by systems that are technologically advanced but responsibly ethical in nature.

To put it succinctly, the advancements of object detection with deep learning are continuously reshaping the computer vision map by introducing new capabilities and opening new potential applications that continue to grow. Challenges solved to understand emerging trends will, of course, push further discoveries into robust efficient and appropriate object detection systems to support modern applications.

9. REFERENCES

[1] Z. Akram, M. Sharif, M. Raza, and S. W. R. Lee, “Automated and computer assisted diagnosis of liver tumor: A review of techniques,” Procedia Comput. Sci., vol. 130, pp. 583-590, 2018.

[2] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, “Deep learning for computer vision: A brief review,” Comput. Intell. Neurosci., vol. 2018, 2018, Art. no. 7068349.

[3] T. Y. Chen, W. Z. Deng, Z. Zhang, and S. W. Guo, “Advances in intelligent video surveillance systems: A comprehensive review,” Tech. Soc. Sci. Eng., vol. 7, no. 4, 2017.