# Evolution of Object Detection and Tracking: From Early Algorithms to Modern Deep Learning Approaches

Sk Babul Akhtar[1,*], Tomal Suvro Sanyashi[1], Tanmay Sinha Roy[1], Shreya Adhikary[1], Debasis Mondal[1]

[1]*Swami Vivekananda University, Barrackpore, 700121*

**Corresponding Author*

**Abstract:** This paper presents a comprehensive review of the evolution of object detection and tracking, tracing the trajectory from early algorithmic approaches to modern deep learning techniques. The historical progression is examined, beginning with foundational methods such as edge detection, template matching, and classical tracking algorithms like Kalman filters and optical flow. The paper then explores the paradigm shift brought by the advent of deep learning, highlighting the impact of Convolutional Neural Networks (CNNs) on object detection through models like R-CNN, YOLO, and SSD. Similarly, advances in object tracking are discussed, focusing on deep learning-based frameworks such as Siamese networks and the integration of transformers. Key challenges, including real-time processing, accuracy, and ethical considerations, are also addressed. The paper concludes by identifying emerging trends and potential future directions, emphasizing the ongoing innovation in this critical area of computer vision.

## Introduction

Object detection (Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. 2023) and tracking are fundamental components of computer vision, enabling systems to perceive and understand the visual world by identifying and following objects (Zhao, Z. Q., Zheng, P., Xu, S. T., & Wu, X. 2019) within a scene. These tasks are crucial across a wide range of applications, from autonomous vehicles and surveillance systems to robotics and augmented reality. Object detection (Papageorgiou, C., & Poggio, T. 2000) involves identifying instances of semantic objects of a certain class within an image, while tracking (Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., & Kim, T. K. 2021) focuses on the continuous localization of these objects across a sequence of frames. The significance of these processes in computer vision lies in their ability to bridge the gap between raw pixel data and higher-level scene interpretation, enabling machines to interact with their environment in a meaningful way.

The objective of this paper is to provide a detailed review of the evolution of object detection and tracking methodologies, spanning from the early algorithmic techniques to the cutting-edge deep learning approaches that dominate the field today. By examining the historical developments, the paper aims to elucidate the transition from classical methods, such as edge detection and template matching, to modern frameworks powered by convolutional neural networks (CNNs) (Chua, L. O. 1997) and transformers. The scope of this review encompasses both the foundational algorithms that laid the groundwork for current technologies and the latest advancements that have pushed the boundaries of accuracy and efficiency in real-world applications. Through this exploration, the paper seeks to offer insights into the ongoing challenges and potential future directions in the domain of object detection and tracking, emphasizing their pivotal role in advancing computer vision.

## Historical Background

The early beginnings of object detection were rooted in fundamental image processing techniques like edge detection and feature extraction, with methods such as the Sobel operator (Tomasi, C. 2012) and Histogram of Oriented Gradients (HOG) (Dalal, N., & Triggs, B. 2005, June) enabling basic object localization. Template matching and simple classifiers like Support Vector Machines (SVM) were later used to detect specific object categories. Object tracking evolved concurrently, with early techniques relying on correlation filters, optical flow, and centroid-based tracking (e.g., Mean Shift) (Carreira-Perpinán, M. A. 2015) to follow objects over time. Key milestones include the development of the Viola-Jones algorithm for real-time face detection, the introduction of Kalman and particle filters for tracking, and the breakthrough of convolutional neural networks (CNNs) (Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. 2021), which revolutionized both detection and tracking with models like R-CNN (Bharati, P., & Pramanik, A. 2020) and Siamese networks (Lu, X., Li, B., Yue, Y., Li, Q., & Yan, J. 2019). These advancements paved the way for more sophisticated, real-time, and highly accurate systems.

## Classical Object Detection and Tracking Techniques

Edge detection and feature extraction have been foundational techniques in object detection, serving as the initial steps in identifying objects within an image by highlighting boundaries and significant regions. The Sobel and Canny edge detectors are classic algorithms that compute the gradient of pixel intensity to detect edges, effectively outlining objects in a scene.

The Histogram of Oriented Gradients (HOG) further advanced feature extraction by capturing local gradient orientations, which are then used to construct a robust representation of object shapes, making it particularly effective for tasks such as pedestrian detection. Template matching (Brunelli, R. 2009) is another early approach where pre-defined patterns or templates of the object are slid across the image to identify regions with high similarity. This method is computationally simple but limited by its sensitivity to variations in object scale, orientation, and appearance. Traditional machine learning approaches marked a significant step forward by introducing more flexible and data-driven methods. Support Vector Machines (SVMs) (Steinwart, I., & Christmann, A. 2008) were employed to create decision boundaries that separate objects from the background based on extracted features. Decision trees and boosting methods like AdaBoost enhanced this by combining weak classifiers to form a stronger, more accurate model. These approaches laid the groundwork for more sophisticated classifiers that could generalize better across different object categories. Cascade classifiers, particularly the Viola-Jones algorithm (Wang, Y. Q. 2014), revolutionized real-time object detection, especially for face detection. This method uses Haar-like features (Lienhart, R., & Maydt, J. 2002, September) to represent objects and constructs a cascade of increasingly complex classifiers that quickly eliminate negative regions while focusing computational resources on more promising areas. The efficiency of this approach made it feasible to perform object detection in real-time, even on limited hardware.

In the realm of tracking, correlation-based methods are fundamental, where correlation filters are used to track objects by matching a template of the object from one frame to subsequent frames. This method, often referred to as "tracking by detection," (Andriluka, M., Roth, S., & Schiele, B. 2008, June) relies on continuously updating the template as the object (Kalal, Z., Mikolajczyk, K., & Matas, J. 2011) moves, maintaining accuracy over time. Optical flow methods, such as the Lucas-Kanade (Oron, S., Bar-Hille, A., & Avidan, S. 2014) and Horn-Schunck (Bruhn, A., Weickert, J., & Schnörr, C. 2005) techniques, estimate the motion of objects by analyzing changes in pixel intensity between consecutive frames. These methods are particularly effective for capturing small, continuous movements, making them suitable for tasks like tracking the motion of individuals in a video sequence. The Kalman filter (Simon, D. 2001), and its more sophisticated counterpart, the particle filter, are widely used for tracking objects in noisy environments. The Kalman filter (Khodarahmi, M., & Maihami, V. 2023) provides a recursive solution to estimate the state of a moving object based on a series of measurements, assuming linear motion and Gaussian noise. The particle filter extends this to handle non-linear motion and non-Gaussian noise by representing the state with a set of weighted particles, each representing a possible state of the object. Finally, the Mean Shift and CAMShift (Exner, D., Bruns, E., Kurz, D., Grundhöfer, A., & Bimber, O. 2010, June) algorithms are centroid-based tracking methods that iteratively converge on the densest region of data points, effectively tracking objects by shifting the centroid towards the peak of a distribution. These algorithms are particularly effective for tracking objects that undergo significant changes in scale and orientation, as they dynamically adjust the search window to follow the object's movement.

These techniques (Forsyth, D. A., & Ponce, J. 2003), spanning from edge detection to advanced tracking methods, represent the foundational tools that have enabled the development of more complex and accurate object detection and tracking systems in modern computer vision.

**Transitioning into Modern Object Tracking Techniques**

The introduction of deep learning (Prince, S. J. 2012) into the realm of object detection and tracking marked a paradigm shift, revolutionizing the field with unprecedented levels of accuracy and robustness. At the core of this transformation lies the development and application of Convolutional Neural Networks (CNNs), which have become the backbone of modern computer vision systems. CNNs introduced a hierarchical approach to feature extraction, allowing the network to automatically learn and refine features across multiple layers, each layer capturing increasingly complex patterns from the raw pixel data. This shift from manually engineered features to deep learning-based feature learning enabled object detection systems to generalize better across diverse datasets and perform well even in complex and cluttered environments.

Early applications of deep learning in object detection were exemplified by the pioneering work on Regions with Convolutional Neural Networks (R-CNN). R-CNN introduced a two-stage process: first, generating region proposals that could potentially contain objects, and second, applying a CNN to classify these regions. This approach significantly improved detection accuracy but was computationally expensive due to the need to run a CNN on each region proposal. Fast R-CNN (Girshick, R. 2015) addressed this bottleneck by introducing a shared computation strategy, where a single forward pass of the CNN (Ren, S., He, K., Girshick, R., & Sun, J. 2016) processed the entire image, followed by region-specific computations using Region of Interest (RoI) pooling. Faster R-CNN further optimized this process by integrating the region proposal generation directly into the network, using a Region Proposal Network

(RPN) (Tang, P., Wang, X., Wang, A., Yan, Y., Liu, W., Huang, J., & Yuille, A. 2018) that dramatically accelerated detection speed while maintaining high accuracy. Single Shot Detectors (SSDs) (Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. 2016) and the YOLO (You Only Look Once) (Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. 2022) family of models represent a significant leap towards real-time object detection by abandoning the region proposal stage entirely. These models predict object classes and bounding boxes directly from feature maps in a single pass, hence the term "single shot." SSD achieves this by utilizing a series of convolutional layers (Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., & Liang, Z. 2019) at different scales, each responsible for detecting objects of varying sizes. YOLO, on the other hand, divides the image into a grid and predicts bounding boxes and class probabilities for each grid cell, allowing it to perform detection at a remarkable speed. These approaches, particularly YOLO, have become synonymous with real-time object detection, being widely adopted in applications where speed is critical, such as autonomous driving and real-time video analysis. Region-based approaches like the Region-based Fully Convolutional Networks (R-FCN) (Dai, J., Li, Y., He, K., & Sun, J. 2016) further refined the detection pipeline by combining the strengths of fully convolutional networks with region-based methods. R-FCN (Li, Z., Chen, Y., Yu, G., & Deng, Y. 2018, April) introduced a position-sensitive score map, where the network outputs a set of score maps for each position in the grid, allowing for precise localization of objects within the proposed regions. This technique maintained the high accuracy of region-based methods while offering improved computational efficiency, making it suitable for large-scale applications.

The emergence of anchor-free detectors like CenterNet (Xu, Z., Hrustic, E., & Vivet, D. 2020), CornerNet (Law, H., Teng, Y., Russakovsky, O., & Deng, J. 2019), and Fully Convolutional One-Stage Object Detection (FCOS) (Tian, Z., Chu, X., Wang, X., Wei, X., & Shen, C. 2022) represents another innovative shift in object detection. These models eliminate the need for predefined anchor boxes, which are traditionally used to predict bounding boxes. Instead, they predict keypoints, such as object centers (CenterNet) or corners (CornerNet), directly from the feature maps, enabling more flexible and accurate detection, especially for objects of varying shapes and sizes. FCOS, in particular, introduced a per-pixel prediction mechanism (Tian, Z., Shen, C., Chen, H., & He, T. 2020) that simplifies the detection pipeline, allowing the network to predict object locations without relying on anchor boxes, (Wang, N., Gao, Y., Chen, H., Wang, P., Tian, Z., Shen, C., & Zhang, Y. 2020) thus reducing the computational burden and improving detection efficiency.

In parallel with advancements in detection, deep learning (Li, Z., & Xu, J. 2021) has also profoundly impacted object tracking, particularly through the introduction of Siamese networks, GOTURN, and MDNet (Zhang, Z., Xie, Y., Xing, F., McGough, M., & Yang, L. 2017). Siamese networks, for instance, are designed to compare the similarity between a template image of the target object and subsequent frames, enabling robust tracking even under challenging conditions like occlusions or background clutter. GOTURN (Generic Object Tracking Using Regression Networks) uses a CNN to directly regress the position of the object in each frame, offering a simple yet effective approach to tracking. MDNet (Multi-Domain Network) takes this further by incorporating domain-specific information during training, allowing the tracker to adapt to different types of objects and scenes, thereby improving generalization across various tracking scenarios. The tracking-by-detection paradigm leverages powerful object detectors within the tracking framework, where an object is first detected in the initial frame and subsequently tracked by repeatedly applying the detector in successive frames. This approach benefits from the accuracy of modern detectors like Faster R-CNN and SSD, providing a robust method for tracking objects that might undergo significant appearance changes over time. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks (Staudemeyer, R. C., & Morris, E. R. 2019) have also found applications in temporal tracking, where the goal is to model the sequential nature of video data. By capturing the temporal dependencies between frames, RNNs (Salehinejad, H., Sankar, S., Barfett, J., Colak, E., & Valaee, S. 2017) and LSTMs can predict the future positions of objects based on their past movements, making them particularly useful for tracking objects in dynamic and unpredictable environments. The latest advancements in object tracking have seen the introduction of transformers, specifically Vision Transformers (ViTs), (Zhai, X., Kolesnikov, A., Houlsby, N., & Beyer, L. 2022) which have demonstrated remarkable success in various computer vision tasks, including tracking. Transformers, with their attention mechanisms, excel at capturing long-range dependencies and complex interactions within the data, making them well-suited for tracking tasks that require a high degree of precision and adaptability. The application of transformers in tracking has opened new avenues for research, pushing the boundaries of what is possible in real-time, high-accuracy object tracking.

## Applications

In autonomous vehicles (Hnewa, M., & Radha, H. 2020), object detection and tracking systems are crucial for perceiving and interpreting the vehicle's surroundings, enabling features such as lane-keeping, collision avoidance, and adaptive cruise control. Advanced algorithms process real-time data from cameras and sensors to identify pedestrians, other vehicles, and road obstacles, facilitating safe and efficient navigation. In surveillance and security (Abba, S., Bizi, A. M., Lee, J. A., Bakouri, S., & Crespo, M. L. 2024), object detection and tracking are employed to monitor and analyze video feeds (Varma, S., & Sreeraj, M. 2013, December) from cameras, detecting suspicious activities, and tracking individuals or objects of interest. These systems enhance security by providing real-time alerts and enabling post-event analysis to support law enforcement and forensic investigations. In healthcare and medical imaging (Ganatra, N. 2021, March), object detection techniques are used to analyze medical scans, such as MRI and CT images, to identify and classify anomalies like tumors (Elakkiya, R., Subramaniyaswamy, V., Vijayakumar, V., & Mahanti, A. 2021) or fractures. Automated systems assist radiologists by providing accurate and timely diagnoses, improving patient outcomes and streamlining diagnostic workflows. In robotics (Farag, M., Abd Ghafar, A. N., & ALSIBAI, M. H. 2019, June) and industrial automation, object detection and tracking enable robots to interact with and manipulate objects within a workspace. This includes tasks such as quality inspection, assembly, and material handling, where precise object localization (Maiettini, E., Pasquale, G., Rosasco, L., & Natale, L. 2020) and tracking are essential for optimizing performance and ensuring operational efficiency.

## Conclusion and Future Work

The advancements in object detection and tracking driven by deep learning have significantly transformed the field, yet several critical challenges remain. Real-time performance is a paramount concern, particularly in applications requiring instantaneous responses, such as autonomous driving and surveillance. Despite significant improvements, achieving both high accuracy and real-time processing remains a delicate balance. High-performing models often involve complex architectures that can be computationally intensive, demanding substantial hardware resources and leading to trade-offs between accuracy and speed. As models grow in complexity, ensuring that they operate within the constraints of real-time systems becomes increasingly challenging. Another challenge is generalization to unseen data. While deep learning models, particularly those based on CNNs and transformers, have shown remarkable performance on known datasets, their ability to generalize to new, unseen scenarios remains a concern. Variability in environmental conditions, object appearances, and contexts can impact the robustness of these models. Ongoing research aims to enhance model adaptability and reduce overfitting by incorporating more diverse training datasets and leveraging advanced techniques such as domain adaptation and transfer learning. Ethical considerations and bias are also crucial aspects of modern object detection and tracking systems. The reliance on large-scale datasets, which may contain inherent biases, can lead to discriminatory outcomes or unintended consequences. Ensuring fairness and mitigating biases require continuous scrutiny of training data and the development of algorithms that promote equitable performance across different demographic groups. Furthermore, ethical considerations extend to privacy concerns, especially in surveillance and security applications, where the deployment of detection and tracking technologies must balance efficacy with individual privacy rights.

Future research in object detection and tracking is poised to address several key areas of development. Advances in hardware for real-time processing, such as specialized accelerators and neuromorphic computing devices (Liu, X., Mao, M., Liu, B., Li, H., Chen, Y., Li, B., ... & Yang, J. 2015, June), are expected to enhance the efficiency of deep learning models, enabling more complex and accurate systems to operate in real-time. These hardware improvements will likely play a crucial role in bridging the gap between high-performance algorithms and practical deployment constraints. Integration with 3D object detection (Xu, Q., Zhong, Y., & Neumann, U. 2022, June) is another promising direction. While current models excel in 2D scenarios, extending detection and tracking capabilities to three-dimensional spaces can significantly improve object recognition and interaction in complex environments. This integration is vital for applications like autonomous driving and robotics, where understanding the spatial relationships between objects is crucial. Cross-modal object detection (Zeng, Y., Ma, C., Zhu, M., Fan, Z., & Yang, X. 2021, September) and tracking, which involves combining data from multiple sensor modalities (e.g., RGB, depth, and infrared), offers the potential for more robust and comprehensive scene understanding. By leveraging complementary information from different sensors, systems can achieve improved accuracy and resilience in diverse conditions, such as low light or adverse weather. Edge computing (Abreha, H. G., Hayajneh, M., & Serhani, M. A. 2022) and federated learning represent significant advancements in decentralized processing and data privacy. Edge computing enables the deployment of detection and tracking models on local devices, reducing latency and bandwidth requirements. Federated learning allows for collaborative model training across multiple devices while preserving data privacy, addressing concerns related to data security and privacy in sensitive applications. Finally, the potential of quantum computing in object detection holds the promise of revolutionizing the field. Quantum algorithms (Meedinti, G. N., Srirekha, K. S., & Delhibabu, R. 2023) could potentially address the computational challenges associated with large-scale data processing and complex model training, providing new avenues for improving performance and efficiency. As quantum technology (Li, J., & Ghosh, S. 2020, August) advances, its integration into object detection and tracking could offer breakthroughs in both speed and capability, opening new frontiers (Baek, H., Kim, D., & Kim, J. 2023) for research and application.

## REFERENCES

Abba, S., Bizi, A. M., Lee, J. A., Bakouri, S., & Crespo, M. L. (2024). Real-time object detection, tracking, and monitoring framework for security surveillance systems. Heliyon, 10(15).

Abreha, H. G., Hayajneh, M., & Serhani, M. A. (2022). Federated learning in edge computing: A systematic survey. Sensors, 22(2), 450.

Andriluka, M., Roth, S., & Schiele, B. (2008, June). People-tracking-by-detection and people-detection-by-tracking. In 2008 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1-8). IEEE.

Baek, H., Kim, D., & Kim, J. (2023). Fast Quantum Convolutional Neural Networks for Low-Complexity Object Detection in Autonomous Driving Applications. arXiv preprint arXiv:2401.01370.

Bharati, P., & Pramanik, A. (2020). Deep learning techniques—R-CNN to mask R-CNN: A survey. Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019, 657-668.

Brunelli, R. (2009). Template matching techniques in computer vision: Theory and practice. John Wiley & Sons.

Bruhn, A., Weickert, J., & Schnörr, C. (2005). Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. International Journal of Computer Vision, 61, 211-231.

Carreira-Perpinán, M. A. (2015). A review of mean-shift algorithms for clustering. arXiv preprint arXiv:1503.00687.

Challa, S. (2011). Fundamentals of object tracking. Cambridge University Press.

Chua, L. O. (1997). CNN: A vision of complexity. International Journal of Bifurcation and Chaos, 7(10), 2219-2425.

Comaniciu, D., & Meer, P. (1999, September). Mean shift analysis and applications. In Proceedings of the Seventh IEEE International Conference on Computer Vision (Vol. 2, pp. 1197-1203). IEEE.

Dai, J., Li, Y., He, K., & Sun, J. (2016). R-FCN: Object detection via region-based fully convolutional networks. Advances in Neural Information Processing Systems, 29.

Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) (Vol. 1, pp. 886-893). IEEE.

Elakkiya, R., Subramaniyaswamy, V., Vijayakumar, V., & Mahanti, A. (2021). Cervical cancer diagnostics healthcare system using hybrid object detection adversarial networks. IEEE Journal of Biomedical and Health Informatics, 26(4), 1464-1471.

Exner, D., Bruns, E., Kurz, D., Grundhöfer, A., & Bimber, O. (2010, June). Fast and robust CAMShift tracking. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops (pp. 9-16). IEEE.

Farag, M., Abd Ghafar, A. N., & ALSIBAI, M. H. (2019, June). Real-time robotic grasping and localization using deep learning-based object detection technique. In 2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS) (pp. 139-144). IEEE.

Forsyth, D. A., & Ponce, J. (2003). A modern approach. Computer Vision: A Modern Approach, 17, 21-48.

Ganatra, N. (2021, March). A comprehensive study of applying object detection methods for medical image analysis. In 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 821-826). IEEE.

Girshick, R. (2015). Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1440-1448).

Hnewa, M., & Radha, H. (2020). Object detection under rainy conditions for autonomous vehicles: A review of state-of-the-art and emerging techniques. IEEE Signal Processing Magazine, 38(1), 53-67.

Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2022). A Review of YOLO Algorithm Developments. Procedia Computer Science, 199, 1066-1073.

Kalal, Z., Mikolajczyk, K., & Matas, J. (2011). Tracking-learning-detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(7), 1409-1422.

Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing, 173, 24-49.

Khodarahmi, M., & Maihami, V. (2023). A review on Kalman filter models. Archives of Computational Methods in Engineering, 30(1), 727-747.

Law, H., Teng, Y., Russakovsky, O., & Deng, J. (2019). CornerNet-Lite: Efficient keypoint based object detection. arXiv preprint arXiv:1904.08900.

Li, J., & Ghosh, S. (2020, August). Quantum-soft qubo suppression for accurate object detection. In European Conference on Computer Vision (pp. 158-173). Cham: Springer International Publishing.

Li, Z., & Xu, J. (2021). [Retracted] Target Adaptive Tracking Based on GOTURN Algorithm with Convolutional Neural Network and Data Fusion. Computational Intelligence and Neuroscience, 2021(1), 4276860.

Li, Z., Chen, Y., Yu, G., & Deng, Y. (2018, April). R-FCN++: Towards accurate region-based fully convolutional networks for object detection. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 32, No. 1).

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single Shot Multibox Detector. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14 (pp. 21-37). Springer International Publishing.

Liu, X., Mao, M., Liu, B., Li, H., Chen, Y., Li, B., ... & Yang, J. (2015, June). RENO: A high-efficient reconfigurable neuromorphic computing accelerator design. In Proceedings of the 52nd Annual Design Automation Conference (pp. 1-6).

Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., & Kim, T. K. (2021). Multiple object tracking: A literature review. Artificial Intelligence, 293, 103448.

Maiettini, E., Pasquale, G., Rosasco, L., & Natale, L. (2020). On-line object detection: A robotics challenge. Autonomous Robots, 44(5), 739-757.

Meedinti, G. N., Srirekha, K. S., & Delhibabu, R. (2023). A Quantum Convolutional Neural Network Approach for Object Detection and Classification. arXiv preprint arXiv:2307.08204.

Oron, S., Bar-Hille, A., & Avidan, S. (2014). Extended Lucas-Kanade Tracking. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13 (pp. 142-156). Springer International Publishing.

Papageorgiou, C., & Poggio, T. (2000). A trainable system for object detection. International Journal of Computer Vision, 38, 15-33.

Prince, S. J. (2012). Computer Vision: Models, Learning, and Inference. Cambridge University Press.

Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137-1149.

Salehinejad, H., Sankar, S., Barfett, J., Colak, E., & Valaee, S. (2017). Recent advances in recurrent neural networks. arXiv preprint arXiv:1801.01078.

Simon, D. (2001). Kalman filtering. Embedded Systems Programming, 14(6), 72-79.

Steinwart, I., & Christmann, A. (2008). Support Vector Machines. Springer Science & Business Media.

Staudemeyer, R. C., & Morris, E. R. (2019). Understanding LSTM--a tutorial into long short-term memory recurrent neural networks. arXiv preprint arXiv:1909.09586.

Tang, P., Wang, X., Wang, A., Yan, Y., Liu, W., Huang, J., & Yuille, A. (2018). Weakly supervised region proposal network and object detection. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 352-368).

Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., & Liang, Z. (2019). Apple detection during different growth stages in orchards using the improved YOLO-V3 model. Computers and Electronics in Agriculture, 157, 417-426.

Tian, Z., Chu, X., Wang, X., Wei, X., & Shen, C. (2022). Fully convolutional one-stage 3D object detection on lidar range images. Advances in Neural Information Processing Systems, 35, 34899-34911.

Tian, Z., Shen, C., Chen, H., & He, T. (2020). FCOS: A simple and strong anchor-free object detector. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(4), 1922-1933.

Varma, S., & Sreeraj, M. (2013, December). Object detection and classification in surveillance system. In 2013 IEEE Recent Advances in Intelligent Computational Systems (RAICS) (pp. 299-303). IEEE.

Wang, N., Gao, Y., Chen, H., Wang, P., Tian, Z., Shen, C., & Zhang, Y. (2020). NAS-FCOS: Fast neural architecture search for object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 11943-11951).

Wang, Y. Q. (2014). An analysis of the Viola-Jones face detection algorithm. Image Processing On Line, 4, 128-148.

Wu, Y., Lim, J., & Yang, M. H. (2013). Online object tracking: A benchmark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2411-2418).

Xu, Q., Zhong, Y., & Neumann, U. (2022, June). Behind the curtain: Learning occluded shapes for 3D object detection. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 36, No. 3, pp. 2893-2901).

Xu, Z., Hrustic, E., & Vivet, D. (2020). Centernet heatmap propagation for real-time video object detection. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16 (pp. 220-234). Springer International Publishing.

Yilmaz, A., Javed, O., & Shah, M. (2006). Object tracking: A survey. ACM Computing Surveys (CSUR), 38(4), 13-es.

Zhang, Z., Xie, Y., Xing, F., McGough, M., & Yang, L. (2017). MDNet: A semantically and visually interpretable medical image diagnosis network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 6428-6436).

Zeng, Y., Ma, C., Zhu, M., Fan, Z., & Yang, X. (2021, September). Cross-modal 3D object detection and tracking for auto-driving. In 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 3850-3857). IEEE.

Zhao, Z. Q., Zheng, P., Xu, S. T., & Wu, X. (2019). Object detection with deep learning: A review. IEEE Transactions on Neural Networks and Learning Systems, 30(11), 3212-3232.

Zhai, X., Kolesnikov, A., Houlsby, N., & Beyer, L. (2022). Scaling vision transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 12104-12113).

Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. Proceedings of the IEEE, 111(3), 257-276.