iJRASET
International Journal For Research in
Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 14 **Issue:** II **Month of publication:** February 2026

**DOI:** https://doi.org/10.22214/ijraset.2026.76991

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

# A Comprehensive Review of Real-Time Shape Detection and Object Tracking Techniques in Computer Vision

Sachin Chavan[1], Dr. Ramesh Manza[2]
[1]Research Student, Dr. Babasaheb Ambedkar Marathwada University, Chh. Sambhaji Nagar, India,
[2]Professor, Dr. Babasaheb Ambedkar Marathwada University, Chh. Sambhaji Nagar, India

***Abstract: Real time shape determination and tracking objects are the main components of modern computer vision that lead to the range of high-level uses that include autonomous navigation, robotics, smart surveillance, and human-computer interfaces. Given the growing dependence on autonomous systems, there is an increasingly pressing need to have algorithms that can provide supreme accuracy and computational latency at the same time. The review provides an in-depth discussion of the development of shapes detection and object tracking methodologies and its latest state-of-the-art methods, focusing on real-time operation. It is critical of usage transforming traditional methods of image processing, like edge detection and Kalman filter, to deep learning models, like Convolutional Neural Networks (CNNs) and more recent models like Vision Transformers (ViTs). Moreover, the review groups and analyzes popular open-source datasets, such as COCO, MOTChallenge, and KITTI, and outlines the consequences of owned data sources. Quantitative empirical evaluation of the integrated systems (Tracking-by-Detection and Joint Detection and Embedding (JDE)) is given and supported with the quantitative performance measures like Mean Average Precision (mAP) and Multi-Object Tracking Accuracy (MOTA). Lastly, the paper establishes the enduring issues of occlusion, variation of scales, deployment of edges, and suggests future research avenues that could help solve the gap between theory and practical applications.***
***Keywords: Object Tracking, Shape Detection, Contour Analysis, Edge Detection, Object Detection, etc.***

## I. INTRODUCTION

Shape detection and tracking of objects are the spatial and temporal aspect of computer vision, respectively. Shape detection detects a geometric structure and object localities in a picture, and it can develop to instance segmentation within the more recent deep-learning settings. Object tracking can be defined as an approximation of the state of a target object in a scene in successive video frames and thus maintaining a distinct identity in spite of motion or concealment [1]. The first requirement, object detection, involves localizing the objects and placing them within bounding boxes as well as classifying them to predefined categories. Although detection solves the query what and where, tracking the query where it is going, and shape detection query what is its precise form [2]. The importance of such technologies has increased many times due to Industry 4.0 and Autonomous systems. During autonomous driving, the vehicle should not just identify a pedestrian and but also their precise body posture through shape analysis to determine movement direction, and from then on, their route as time progresses to avoid collisions. A comparable case is in medical imaging where tracking of instruments used in surgery is done in real-time, with accuracy of the shape of the organs, is required in robotic-assisted surgery. The ability to handle this information at real-time, generally above 30 FPS is the limiting factor that draws a line between theory and applications of practical use [3][4]. The sphere still has to face the challenges that persist even with the rapid progress. Its environmental effects which include changes in illumination, motion blur, and background clutter are a problem to shape extraction [5]. Switching identities is a highly sensitive issue when it comes to tracking; there are also identity switches caused by occlusion, and the computational cost of using heavy, frame-based detectors, they will also still serve as points of weakness. Furthermore, the use of huge annotated datasets readily creates bias and restricts the applicability to new fields [5].

The contribution and study of this research article are the following:

1) An overview of different articles associated with shape detection and object tracking, through the history of turning a human portion characterized by hand procedures into transformer-based constructions.
2) An overview of a multitude of datasets, namely, the examination of utility and domain-specificity of open-source benchmarks and proprietary data.

3) An overview of various techniques, which are in terms of the algorithmic task of object detection, shape segmentation, and temporal tracking. Measurement parameters applied in the process of measuring performance during detection and tracking.

## II. LITERATURE REVIEW

The computer vision development can be divided into the period of manual features and the period of deep learning. In this section, the major developments which define this transition are reviewed. Initial studies in shape recognition strongly depended on low level image processing. This was the original algorithm used by canny with gradient magnitude, which was used to outline boundaries. Being computationally efficient had no semantic understanding and treated a shadow and an object edge as the same [6]. To solve the shape continuity Kass et al. proposed the use of Snakes or Active Contour Models that minimized they energy functional to fit a spline to object edges. These methods were however sensitive to initialized and noisy [7][8]. The ViolaJones framework used Haar cascades to detect faces, with the use of integral images to enable real-time operation, but could not perform effectively in the protocol of more complicated multi-class detection. Pedestrian detection was subsequently improved by Dalal and Triggs with Histograms of Oriented Gradients [HOG] and Support Vector Machines [SVMs] to create a powerful feature description that was pre-eminent in the pre-deep-learning age [9], [10].

The development of deep learning changed the paradigm. Girshick et al. came up with R-CNN, which applied Convolutional Neural Networks in the detection of objects, much better than HOG-based approaches [11]. Nevertheless, R- CNN was sluggish, as the computations were done repeatedly. It is this limitation that led to Fast R-CNN and later Faster R-CNN, which proposed the use of the Region Proposal Network that allowed near-real-time detection of an image, sharing the convolutional features in the process [12], [13]. At the same time, Redmon et al. put the entire discipline on its heels with YOLO [You Only Look Once], a one-stage detector which solved the regression problem of detection, at the expense of a small sacrifice in accuracy versus two-stage detectors [14]. Liu et al. developed the Single Shot Multibox Detector (SSD) that generally enhanced YOLO by using multi-scale feature maps that enhance the detection of small objects [15], [16].

The paradigm of Tracking-by-Detection is the one that became predominant in tracking. Bewley et al. proposed SORT (Simple Online and Real Time Tracking) that used Kalman filters and Hungarian algorithm to do data association on detection [17]. It was further expanded by Wojke et al. to be incorporated into DeepSORT, a method that uses an appearance-based CNN-based Re-Identification [Re-ID] descriptor to minimize switches of identity during occlusions [18]. Recent research directions appear to treat Joint Detection and Embedding [JDE] whereby in these models detection and appearance embedding are learned concurrently to save on computational cost [19]. Lastly, Transformer has spawned DETR and TrackFormer, repositioning tracking as a set prediction task with attention mechanisms, creating a new direction that eliminates manual components like non maximum suppression [20], [21], [22].

## III. DATASETS

The modern computer-vision algorithms run on data. The performance of the trained models is determined by the quality, diversity, and precision of the annotation of the datasets.

### A. Open-source datasets

Open-source datasets offer a universal point of reference to the scholarly community.

1) *COCO [Common Objects in Context]:* It is an object detector and instance segmentation benchmark. It has more than 200,000 labeled images containing 80 object categories. They are bounding boxes and pixel-wise segmentation masks annotated, which is essential in the evaluation of shape detection [23].
2) *PASCAL VOC*: The PASCAL VOC is an earlier benchmark with smaller scale than COCO, but with vital importance in comparison with the past. It presents standardized image datasets of object class recognition and it contains 20 object categories that have bounding box and segmentation annotations [24].
3) *Open Images Dataset:* This is a dataset generated by Google, which is large, with about 9 million images, which are supplemented with bound-box annotations, object segmentation, and visual relations and is based on an even broader set of classes compared to COCO [25].
4) *KITTI:* A virtual environment on autonomous driving. It offers video streams together with stereo, optical flow, and 3D object-detection labels (cars, pedestrians, cyclists) and it is customary to use it as a 3D shape detection and tracking benchmark [26].
5) *MOTChallenge (MOT15, MOT16, MOT20):* It is the current benchmark of Multi-Object Tracking (MOT). It pays attention to tracking people in a busy place, and it gives video sequences with demanding cases of occlusion [27].

6) *DAVIS (Densely Annotated Video Segmentation):* This algorithm is specially made to annotate video objects. It also generates pixel-perfect masks and masks of the video in each frame, which is the reference point of accuracy of the shape tracking [28].
7) *ShapeNet*: One of the datasets on 3D shape, with much CAD models. Algorithms that seek to establish a 3D geometry based on 2 -D images are necessary. Cityscapes: Studies semantic perception of city streets. It offers annotations of pixel-level semantic and instance segmentation that is important in automotive shape recognition [29].

### *B. Non-Open-Source Datasets*

Proprietary data sets are usually found in sensitive areas or any commercial entity. Hospitals in the medical world have enormous databases of annotated MRI and CT scans, which are applied to determine tumor shapes; these are considered confidential based on HIPAA laws and GDPR. Tesla and Waymo companies gather billions of miles of video information in the industry of autonomous vehicles. This information includes unusual edge cases such as extreme weather, unusual accidents which are not represented in publicly available datasets. Although it is better in training strong industrial models, it is problematic in the academic world because it is not accessible to such datasets and results in a lack of reproducibility and independent validation [30], [31].

## IV. METHODS

The approach involved in the identification of forms, the recording of movements as well as the identification of objects are described here with each object classifying based on their core purpose and architectural philosophy.

### *A. Object Shape Detection Methods.*

Geometric analysis has been substituted by the pixel-by-pixel classification. Classical Geometric Methods Simple geometric primitive's lines and circles are identified by a method based on the Hough Transform, which uses a voting algorithm on parameter space. Active Contours (Snakes) adjust a curve over time in order to follow strong gradients giving a mathematical representation of the shape. Instance Segmentation Deep Learning is the newest paradigm of detecting shapes. Mask R -CNN works on the principle of a mask branch addition to the Faster R -CNN framework. It employs a Region of Interest Align [RoIAlign] layer to retain the precise spatial locations, which anticipates a binary mask of each object instance [13]. Real-Time Segmentation YOLACT searches in a smaller set of prototypes by-passing the feature localization stage and fuses masks of prototypes with learned coefficients together. This enables it to operate with excess of 30 FPS on average GPUs [16].

### *B. Object Tracking Methods*

1) *Kalman Filtering:* This is a recursive algorithm used to determine the state of a moving object (that is its position and velocity) through minimization of the mean squared error. It is also computationally parsimonious and works well with linear movement [17], [33]. Kalman Tracking also known as Tracking-by-Detection (TBD). It entails the execution of an object detector on each frame and the matching of the boxes. SORT applies the Hungarian algorithm to match data with associations and Kalman filter to make predictions based on Intersection over Union [17], [33].
2) *Appearance-Based Tracking:* DeepSORT is an extension of SORT, which incorporates a deep appearance description. An independent neural network is an extractor of feature vectors of the identified bounding box. The association measure is a variant of motion-based and appearance similarity [Cosine distance] criteria which enables the tracker to re-detect the objects after being lost to view [18], [19].

### *C. Object Detection Methods*

Faster R-CNN model finds few region proposals and classifies them. They are very accurate and their computational cost is prohibitive often with frame rate less than fifteen frames per second [12]. The examples of one-stage detectors include YOLO (You Only Look Once) and SSD, which consider the problem of object detection based on a single regression model. These classifiers do their work as a single regression, in which pixels in an image are directly translated to bounding-box locations, and probabilities of classification [14]. Recent versions, such as YOLOv8 and YOLOv9, use CSPDarknet backbones and PA -FPN necks to obtain state-of-the-art trade-offs between accuracy and speed [14]. Multi-Object Tracking and Segmentation [MOTS] is traditional tracking is based on bounding box which accommodates background noise. MOTS is a paradigm of pixel-based tracking [33]. Extensions of the Mask R-CNN include Track R-CNN which adds 3D convolutions to the Mask R-CNN algorithm, allowing time-driven persistent tracking of the binary mask of the target object but not simply the box. Transformer Based Solutions TrackFormer is a joint detector and tracker that uses track queries. The model operates on the queries of the last frame to the present one through an attention mechanism, implicitly performing association and shape prediction in one end-to-end differentiable network [20], [21].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

## V. COMPARISON ANALYSIS

In order to reach an in-depth knowledge of the existing state of technological nature, we undertake a comparative analysis of ten innovative papers, which specify the course of field development. The discussion evaluates how accuracy-oriented, slower models have changed to real time, efficient structures.

Table 1: Comparative Study of the shape detection key and object tracking algorithm.

| Sr. No. | Reference | Author & Year | Dataset | Method / Technique | Result / Performance |
|---|---|---|---|---|---|
| 1 | [12] | Ren et al., 2015 | PASCAL VOC | Faster R-CNN (Two-Stage Detection) | 73.2% mAP 7 FPS |
| 2 | [14] | Redmon et al., 2016 | PASCAL VOC | YOLOv1 (One-Stage Detection) | 63.4% mAP 45 FPS |
| 3 | [15] | Liu et al., 2016 | COCO | SSD (Single Shot MultiBox) | 41.2% mAP 46 FPS |
| 4 | [13] | He et al., 2017 | COCO | Mask R-CNN (Instance Seg.) | 37.1% AP (Mask) 5 FPS |
| 5 | [17] | Bewley et al., 2016 | MOT15 | SORT (Tracking-by-Detection) | 33.4% MOTA 260 FPS |
| 6 | [18] | Wojke et al., 2017 | MOT16 | DeepSORT (Tracking + ReID) | 61.4% MOTA 40 FPS |
| 7 | [16] | Bolya et al., 2019 | COCO | YOLACT (Real-time Seg.) | 29.8% AP (Mask) 33 FPS |
| 8 | [19] | Wang et al., 2020 | MOT16 | JDE (Joint Detection Embedding) | 64.4% MOTA 18 FPS |
| 9 | [20] | Carion et al., 2020 | COCO | DETR (Transformer Detection) | 42.0% AP 28 FPS |
| 10 | [34] | Zhang et al., 2022 | MOT17 | ByteTrack (Association Method) | 80.3% MOTA 30 FPS |

The comparative analysis shows that, despite the fact that the two-stage detections are the most commonly used architectures to set the standard of accuracy throughout the history, the industry has mainly adopted the use of YOLO-based architectures in real-time applications. Translocation to track, the paradigm remains persistently popular, which is the Tracking-by-Detection paradigm as witnessed by the case of ByteTrack and DeepSORT due to their modularity and scale-ability. However, more speed-accuracy gap is becoming smaller with the use of transformer-based methods like DETR [34].

## VI. EVALUATION MATRICES

Quantitative performance is a mandatory part of the measuring advances in the area of computer vision.

*1)* Intersection over Union (IoU): This is a basic metric which is used in both detection and shape accuracy. It determines the geometrical penetration of the predictive area (as a localization box or a mask) and the ground truth.

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union}$$

An approximation of a prediction is usually counted in a True Positive case when the IoU is greater than some set value (such as 0.5) [35].

*2)* Mean Average Precision (mAP): This metric is one canonical object-detection metric which is the mean of the area under the Precision, Recall curve across individual object classes. The COCO benchmark is the traditional performance metric that presents the mean Average Precision [mAP] at different IoU thresholds (0.5 to 0.95) and averages the results across numerous different thresholds to cancel the effects of poor localizations [35].

3) Multiple Object Tracking Accuracy (MOTA): The MOTA is the main tracking measure that integrates three error types false positives (FP), targets not detected (FN) and identity switches [IDSW]

$$MOTA = 1 - \frac{\sum(FN_t + FP_t + IDSW_t)}{\sum GT_t}$$

In spite of the fact that MOTA is useful in the measurement of the general tracking accuracy, it does not punish identity switches that much.

4) IDF1 Score: This measure is used to represent the harmonic mean of the correctly recognized detections at the average number of ground-truth and computed detections, meaning that increasing the average number of detections in ground-truth and computed data sets does not increase the IDF1 score. In contrast to MOTA, IDF1 focuses on the time interval during which the tracker retains the correct identity and it makes it a more appropriate measure of tracking stability [35].
5) Frames Per Second (FPS): This is a throughput measure that determines real time capacity, when the application demands high performance, typically the value should be more than 30 FPS [35].

## VII. CONCLUSION

This review gives a methodological review on real-time shape detection and tracking of objects to outline the path traversed by heuristically based edge detection and complex deep-learning based pipelines. We show that on limited datasets that may be considered as a static benchmark, e.g. COCO, the accuracy has outdone ethical human performance; however, maintaining accuracy on demanding datasets like real-time video processing is quite challenging. The current superiority of the YOLO family in detection, as well as the hybrid trackers such as DeepSORT, confirm the desire of the community to find the optimal balance between both speed and accuracy. Also, there is a achieved concept of shape detection into tracking, as in the andragogy of the MOTS framework, that provides a deeper understanding of the scene, yet with a high computational cost. Lastly, implementation of Vision Transformers promises to change the paradigm to more globally, context-aware reasoning, but still requires additional optimization to achieve the performance of convolutional neural network on edge devices.

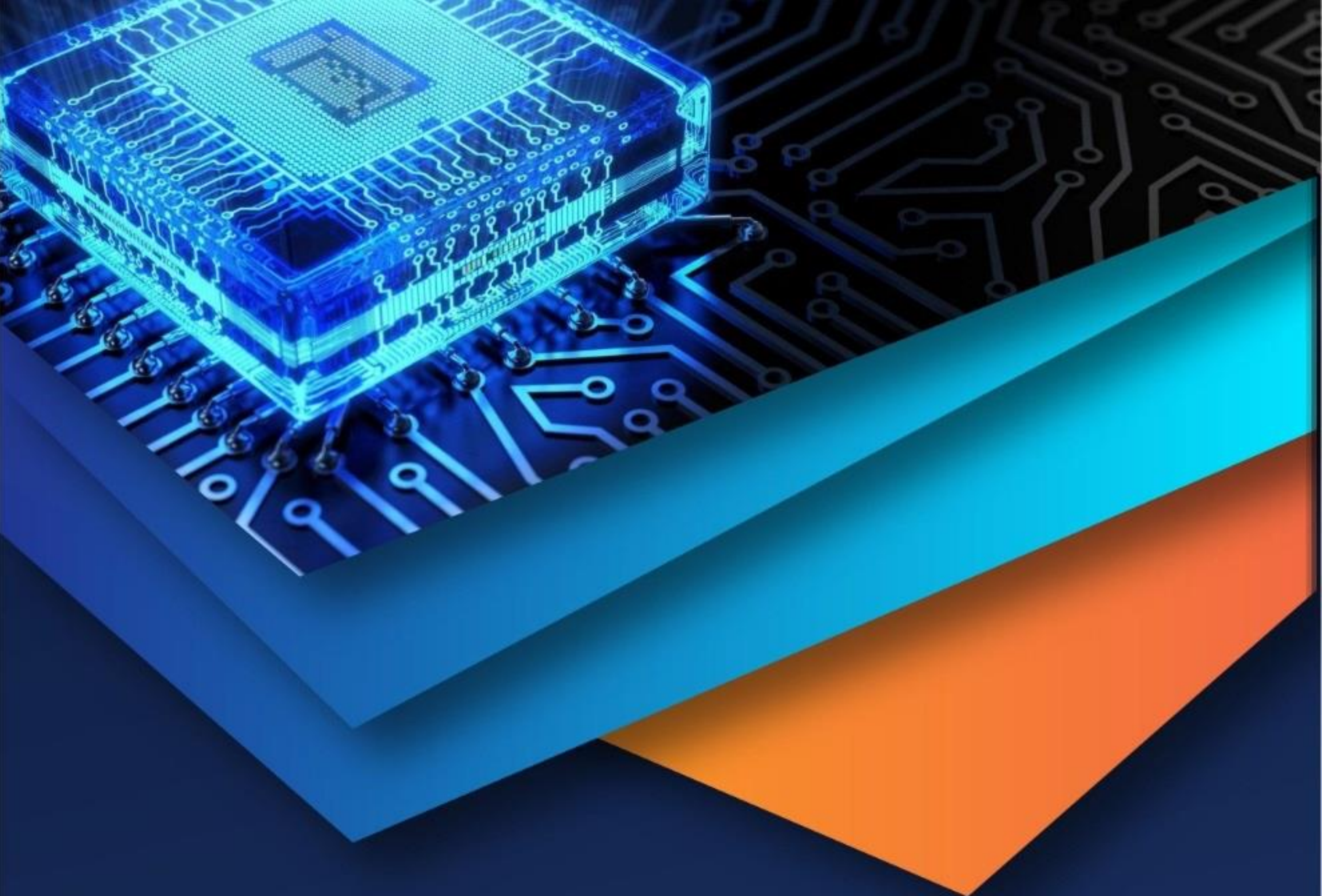
## VIII. FUTURE RESEARCH DIRECTIONS

End-to-End Unsupervised Learning: The existing approaches are based on annotated data to a significant extent. Future studies ought to focus on self-supervised methodologies, including Masked Auto encoders, that are capable of acquiring consistent object representations and geometries through observation of unlabeled video streams, thus mimicking the human visual perception developmental processes. Edge AI and TinyML: The research has to make a shift towards model compression, quantization, and neural architecture search (NAS) to reduce floating-point operations (FLOPs) and avoid losing much accuracy. Multi-Mode Fusion: The utilization of RGB data is not sufficient in different weather conditions. Sensor fusion of LiDAR and thermal cameras signals via radar will be instrumental in order to have resilient tracking under autonomous driving conditions. Long-term Occlusion Processing: Existing trackers have weaknesses in being used to track objects that spend long periods in the occlusion. Graph Neural Networks (GNNs) is an emerging solution to consider long-term time-related dependencies and inter-object relationships, which will solve the problem of re-identification.

## REFERENCES

[1] Wagh, C. (2023). Object detection and tracking using deep learning and OpenCV in real time environment. Int J Eng Res Technol (IJERT), 12(04).
[2] Alkhamaiseh, K. N., Grantner, J. L., Abdel-Qader, I., & Shebrain, S. (2023). Towards real-time multi-class object detection and tracking for the FLS pattern cutting task. Adv. Sci. Technol. Eng. Syst. J., 95(6), 87-95.
[3] Godil, A., Bostelman, R., Shackleford, W., Hong, T., & Shneier, M. (2014). Performance metrics for evaluating object and human detection and tracking systems. National Institute of Standards and Technology, 1-16.
[4] Jiao, L., Zhang, F., Liu, F., Yang, S., Li, L., Feng, Z., & Qu, R. (2019). A survey of deep learning-based object detection. IEEE access, 7, 128837-128868.
[5] Habash, N., Alqumsan, A. A., & Zhou, T. (2025). Recent Real-Time Aerial Object Detection Approaches, Performance, Optimization, and Efficient Design Trends for Onboard Performance: A Survey. Sensors, 25(24), 7563.
[6] Pagire, V., Chavali, M., & Kale, A. (2025). A comprehensive review of object detection with traditional and deep learning methods. Signal Processing, 237, 110075
[7] Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active contour models. International journal of computer vision, 1(4), 321-331
[8] Mirzaei, B., Nezamabadi-Pour, H., Raoof, A., & Derakhshani, R. (2023). Small object detection and tracking: a comprehensive review. Sensors, 23(15), 6887.
[9] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, 2001, pp. I-I, doi: 10.1109/CVPR.2001.990517.

[10] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.
[11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2014.[10]
[12] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.
[13] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).
[14] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.
[15] Liu, W. et al. (2016). SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science(), vol 9905. Springer, Cham. https://doi.org/10.1007/978-3-319-46448-0_2
[16] Bolya, D., Zhou, C., Xiao, F., & Lee, Y. J. (2019). Yolact: Real-time instance segmentation. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 9157-9166).
[17] Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016, September). Simple online and realtime tracking. In 2016 IEEE international conference on image processing (ICIP) (pp. 3464-3468). Ieee.
[18] Wojke, N., Bewley, A., & Paulus, D. (2017, September). Simple online and realtime tracking with a deep association metric. In 2017 IEEE international conference on image processing (ICIP) (pp. 3645-3649). IEEE.
[19] Wang, Z., Zheng, L., Liu, Y., Li, Y., & Wang, S. (2020, August). Towards real-time multi-object tracking. In European conference on computer vision (pp. 107-122). Cham: Springer International Publishing.
[20] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020, August). End-to-end object detection with transformers. In European conference on computer vision (pp. 213-229). Cham: Springer International Publishing.
[21] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020, August). End-to-end object detection with transformers. In European conference on computer vision (pp. 213-229). Cham: Springer International Publishing.
[22] Meinhardt, T., Kirillov, A., Leal-Taixe, L., & Feichtenhofer, C. (2022). Trackformer: Multi-object tracking with transformers. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 8844-8854).
[23] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014, September). Microsoft coco: Common objects in context. In European conference on computer vision (pp. 740-755). Cham: Springer International Publishing.
[24] Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. International journal of computer vision, 88(2), 303-338.A. Kuznetsova et al., "The Open Images Dataset V4," International Journal of Computer Vision, vol. 128, no. 7, pp. 1956–1981, 2020.
[25] Geiger, A., Lenz, P., & Urtasun, R. (2012, June). Are we ready for autonomous driving? the kitti vision benchmark suite. In 2012 IEEE conference on computer vision and pattern recognition (pp. 3354-3361). IEEE.
[26] Geiger, A., Lenz, P., & Urtasun, R. (2012, June). Are we ready for autonomous driving? the kitti vision benchmark suite. In 2012 IEEE conference on computer vision and pattern recognition (pp. 3354-3361). IEEE.
[27] Milan, A., Leal-Taixé, L., Reid, I., Roth, S., & Schindler, K. (2016). MOT16: A benchmark for multi-object tracking. arXiv preprint arXiv:1603.00831.
[28] Perazzi, F., Pont-Tuset, J., McWilliams, B., Van Gool, L., Gross, M., & Sorkine-Hornung, A. (2016). A benchmark dataset and evaluation methodology for video object segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 724-732).
[29] Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., ... & Yu, F. (2015). Shapenet: An information-rich 3d model repository. arXiv preprint arXiv:1512.03012.
[30] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., ... & Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3213-3223).
[31] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical image analysis, 42, 60-88.
[32] C. Badue et al., "Self-driving cars: A survey," Expert Systems with Applications, vol. 165, p. 113816, 2021.
[33] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems.P. Voigtlaender et al., "MOTS: Multi-Object Tracking and Segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019.
[34] Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., ... & Wang, X. (2022, October). Bytetrack: Multi-object tracking by associating every detection box. In European conference on computer vision (pp. 1-21). Cham: Springer Nature Switzerland.
[35] Ristani, E., Solera, F., Zou, R., Cucchiara, R., & Tomasi, C. (2016, October). Performance measures and a data set for multi-target, multi-camera tracking. In European conference on computer vision (pp. 17-35). Cham: Springer International Publishing.




10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429


# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY


Call : 08813907089 (24*7 Support on Whatsapp)