Journal of Al-Turath University College (JATUC)
Open Access Journal
ISSN : 2074-5621
DOI: https://doi.org/10.63964/JATUC.43.1.2026.28
https://jatuc.uoturath.edu.iq/

A Comprehensive Review of Deep Learning Methods for Detection and Tracking of Multiple Objects: A Review

Mohammed S. H. Al-Tamimi*
Computer Science Department, College of Science, University of Baghdad, Baghdad, Iraq

**Correspondence email*: mohammed.s@sc.uobaghdad.edu.iq

KEYWORDS

Deep Learning, Methods, Algorithms, Challenges, Object

ABSTRACT

In recent years, the advancement of deep learning techniques has significantly transformed the fields of computer vision, particularly in object detection and tracking. This paper presents a comprehensive review of state-of-the-art deep learning methods utilized for the detection and tracking of multiple objects in various environments. We categorize the techniques based on their underlying architectures, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid approaches, highlighting their strengths and limitations. The survey covers traditional algorithms as well as emerging deep learning models, examining their performance across different benchmarks and datasets. Furthermore, we address the challenges faced in real-time applications, such as occlusion, scale variation, and computational efficiency. By synthesizing current research trends, this review aims to provide insights for future developments in the domain, guiding researchers and practitioners toward effective solutions for multi-object detection and tracking tasks.

الكلمات المفتاحية

التعلم العميق ,الطرق , الخوارزميات التحديات ,الكائنات .

الملخص

في السنوات الأخيرة، أحدث التقدم في تقنيات التعلم العميق نقلة نوعية في مجال رؤية الحاسوب، لا سيما في مجال اكتشاف وتتبع الأجسام. تقدم هذه الورقة مراجعة شاملة لأحدث أساليب التعلم العميق المستخدمة في اكتشاف وتتبع أجسام متعددة في بيئات متنوعة. نصنف هذه التقنيات بناءً على بنيتها الأساسية، بما في ذلك الشبكات العصبية الالتفافية (CNNs) والشبكات العصبية المتكررة (RNNs) والأساليب الهجينة، مع تسليط الضوء على نقاط قوتها وقيودها. تغطي الدراسة الخوارزميات التقليدية ونماذج التعلم العميق الحديثة، وتفحص أداءها عبر معايير ومجموعات بيانات مختلفة. علاوة على ذلك، نتناول التحديات التي تواجه التطبيقات الآنية، مثل الحجب، وتغير الحجم، والكفاءة الحسابية. من خلال تجميع اتجاهات البحث الحالية، تهدف هذه المراجعة إلى تقديم رؤى للتطورات المستقبلية في هذا المجال، وتوجيه الباحثين والممارسين نحو حلول فعالة لمهام اكتشاف وتتبع الأجسام المتعددة.

1. INTRODUCTION

Multitarget detection and tracking denote the complementary tasks of identifying and following multiple objects of interest in a video sequence. This problem arises in contexts such as autonomous driving, robotic navigation, and surveillance. Several related tasks exist, including detection, tracking, and re-identification; single-frame or per-frame sequences can also be considered[1].

Multitarget detection encompasses the identification and localization of all objects in a frame. Detection methods can be divided into single-shot and two-stage approaches; the former produce detections within a single forward pass, whereas the latter first generate candidate objects for evaluation. Multitarget detection is often treated in a category-agnostic manner; end-to-end systems further minimize the use of intermediate formats [2].

Multitarget tracking estimates the individual trajectories of multiple objects across a video sequence. Independent tracking approaches, which operate frame by frame, break the problem into detection and

successive object management. Tracking-by-detection, applying data associations to a series of detections, remains the most widely adopted paradigm; it estimates identity-preserving associations between detection frames. Joint state estimation approaches model the entire sequence, yet still require detection components. End-to-end methods predict both trajectories and detections from the same input, with or without a tracking-by-detection mechanism [3].

2. BACKGROUND AND DEFINITIONS

Detection and tracking of multiple objects is a research area with practical applications in autonomous driving, traffic analysis, human action analysis for behavior understanding, and multi-object video analysis for many applications, such as intelligent online monitoring and smart city development. Tracking targets across a sequence of frames is a more difficult but more meaningful task than detecting multiple objects in a single image. The task of Multi-Object Tracking (MOT) aims to detect a set of targets and track their motion and trajectories [4].

Video-based multiclass multi-object detection and tracking can be defined as a joint problem where the location of targets is obtained simultaneously, or as a two-stage problem where the motion of targets is tracked based on the detection results. Tracking-by-Detection methods work in a second way, where multi-object detection is performed independently, and data association is resolved subsequently. Continuous Object Detection and Tracking (CODT) is a lightweight pose-aware model that simultaneously detects and tracks thousands of human bodies in real time [5].

Cross-camera tracking requires finishing tracking a target when it leaves a camera and re-tracking it when it appears in another camera, which has spatiotemporal constraints between targets across a long sequence. Crowd object detection is to detect small individuals in extremely crowded situations. Appearance-preserving object detection aims to detect the same object under extreme environmental changes, such as color, size, and rotation [6].

3. MULTITARGET DETECTION: PRINCIPLES AND DEEP LEARNING APPROACHES

Multitarget detection aims to identify and localize all objects of interest in a video frame; it serves as a crucial upstream component for the more challenging task of multitarget tracking. Detection networks can be categorized as single-shot (one-stage) or two-stage frameworks based on their operational principles. Single-shot architectures - including Yolo [7], SSD, and BDD [8] - predict class labels and bounding-box coordinates within a single feed-forward pass through the network. Such architectures remain widely adopted within the computer vision community due to their efficient inference. Multi-Stage Detectors (MSD) that further evolve the region-based detectors in a modular fashion or continue enhancing the overall performance of detectors still follow the two-stage setup. Multitarget detectors employ monitoring schemes that jointly estimate a vehicle's position and its categorization using a vehicle's tracking window sequence spanning K frames, and are conditional. By a previously useful detector framework - a Strongly Supervised Multi-target Object Detection framework. Principles of multitarget detection using deep learning can be summarized into three directions: establishing Measurement functions that unify single-vehicle tracking and category recognition, and defining a Multi-Label Cross Photometric-Metric Space. The functions of the observation processes between true tracking states and measurement observations. A multitarget detection model could provide both prior position for the multitarget tracking task [9].

Generic multitarget detection is enabled through more general low-gap detection methods. The detection hypothesis of all objects in each frame could come from an arbitrary Multi-Object Tracking (MOT) dataset not limited to a single vehicle type; simultaneous Re-identification-Positioning enables detection tasks an trajectory association to remain a challenging and timeless task; it not only gives comprehensive Real-time tracking performance remains an investigate topic especially for multiple category, which is required by autonomous driving system competitive tracking performance;. Appearance cues of time-specific objects assume occupying similar spatial characteristics, and develop techniques that predict sequential index frames by frame [10].

Valid multitarget detection is determined upon valid prior-vehicle tracking, not limited to detectors. Components bleed-sequential tracking, associating pre-chiotank's vehicles feature contained sufficient information, sequentially feeding labelled-hot detection. High-equipment configuration detectors favored acceptance, but were not feasible in common situations. Stable-multi-opacity with plastic deformation

multi-object detection model remains an investigation, and handling large-scale covariance appears tractable on real-world images. The distinct re-identification model generalizes objects across heterogeneous stages. Tracked vehicles generate various coverage-invisible appearance tread spots, which allow weakness prior apparent-object type and pre-estimated vehicle types to still link frame deviating gate-kept unique vehicles in infringements, maintaining real-time performance [11].

4. MULTITARGET TRACKING: FOUNDATIONS AND DEEP LEARNING IMPLEMENTATIONS

Detecting and tracking multiple targets in real-time video surveillance, assuming fixed and known cameras, is challenging in practical applications. The problem is ill-posed in the sense that prior knowledge alone cannot ensure the recovery of the trajectories of each target from the 3D occupancy information. In contrast, the tracking procedure becomes easier when the object-rehydration video is provided, such that the 3D trajectories of the targets or the 3D information are unambiguous. Therefore, the multiple-target tracking algorithm has major significance not only for machine perception but also for society at large. The following solution keeps the strengths of video rehydration to track multiple targets, while ensuring the unambiguous observation of the object. As mentioned, a prior conditional ANFIS is learned when the object is assumed to be detected and known as an SA model. The proposed flow is presented in Fig. 1. Specifically, given an observed sequence of frames and the initial proposal p (t) ={O (t), L (t)} of the target's own bounding boxes and labels at the beginning instant t 0. The first part is to recover the target-hazard scene or object-rehydration frames at the first instant of each target. The set of conditions is shared among other predications in order to represent the target-critical spatial-temporal characteristics. Coverage objectives, such as number and area, help to maintain the topology of the targets. A backward counting-down surrogate of prior AAFs enables filling-unfilled covering along the time dimension. An intermediate set of observations (I, O) of the writings between a given input (I) and an object-own copy (O) can also be determined in only one pass within the rehydration stage. The method selects a small number proportional to the total track length to capture the sketch information of the object [12].

5. CORE ARCHITECTURES FOR DETECTION AND TRACKING

Multitarget detection is the problem of locating, identifying, and counting objects of interest in images, frames, or video sequences. The task of simultaneously carrying out multiple target detection and tracking in the same architecture is being pursued from different perspectives, resulting in different architectural design patterns and inference pipeline structures. In deep learning modular pipelines for multitarget detection followed by tracking, detectors with slow inference times and even some with fast mapping to tracking have been integrated with trackers. Integrated multitarget detection and tracking methods, whose inference times sometimes remain critical for high-throughput applications, still remain a separate line of research. To date, no survey has treated the integrated aspect in a comprehensive manner. A review is presented of specific architectures that have been deployed to address multimodal detection and tracking jointly. A detailed discussion of the relevant components necessary to accomplish detection and tracking is provided. The emphasis is placed on distinguishing different architectural paradigms within the integrated approach, since this permits practitioners to select the family most suitable to their needs, given the current implementation [13].

Two architectures pursue a single-stage concept, where the different stages in other tracking-by-detection approaches are collapsed into a single forward pass. Specifically, proposals are generated within the tracking-and-detection framework, which extends detection approaches. Eleven different architectures-two single-stage, eight two-stage, and one unified-have been identified. Furthermore, several trends in the design of these architectures have also been observed across the surveyed literature. Feature extraction is usually based either on a backbone network or on a combination of a backbone network followed by a feature pyramid, where, in both cases, various options are available within the community. The option of augmenting the detection and tracking tasks through extra information on different modalities, such as depth and optical flow, remains open throughout. Many of the architectures assessed-eight out of the eleven-exploit the temporal dimension directly to enhance the current task of interest by integrating proposals or detections tracked in parallel. Paralleling this strategy, recurrent, transformer, and convolutional flow modeling inserted into the pipeline at different points in the inference phase have also been proposed [1][14].

6. DATA, BENCHMARKS, AND EVALUATION METRICS

Multiple-Object-Tracking (MOT) datasets, benchmarks, and evaluation protocols have been established to streamline the development and independent assessment of detection-and-tracking methods. Comprehensive surveys of contemporary trackers have incorporated these resources, allowing comparative analysis of performance across approximately fifty detections, tracking, and combined detection-and-tracking architectures [15].

A range of publicly accessible datasets with varying modalities, scenarios, and splits has been released to facilitate the evaluation of numerous methods within the field. Reported outcomes across these protocols almost invariably indicate a robust correlation between detection performance and tracking success, underscoring the essentiality of accurate detection for effective tracking, without which temporal and interaction cues cannot be leveraged. Statistical analyses have further unveiled the meaningful relationship between the Mean Average Precision (mAP) metric employed in determining detection accuracy and Multiple-Object-Tracking Accuracy (MOTA) scores, rendering the mAP measure a valuable indicator of tracking capability. No architectural topology has emerged as universally superior for detection-and-tracking applications, although joint, end-to-end, and single-shot approaches exhibit notable efficacy across diverse datasets and are thus particularly highlighted[16].

Cross-dataset generalization constitutes a critical challenge; several competing strategies have been proposed to enable performance transfer from one dataset to another. Proposals to construct an easily extended, relatively unconstrained benchmark that facilitates holistic comparison of object detectors and multi-object trackers across multiple datasets, while mitigating the impact of disparate annotation protocols, continue to be a work in progress. Existing datasets and benchmarks differ widely in terms of object-count distributions (sparcity/density), scenario complexity, annotation granularity (four/ six characteristics), and data modalities, which complicates the establishment of uniform performance baselines and complicates progress tracking toward a globally formulated multi-object-tracking solution.

Eight representative MOT test suites have been compiled, detailing dataset typologies according to activity level and annotation density, while also cataloging the protocol stipulated by each resource. Protocols differ not only in the number of videos but also in protocols, such as being of moderate difficulty or multiple views [17].

7. TEMPORAL MODELING AND MOTION UNDERSTANDING

Multitarget detection and tracking involve learning to foresee and describe future spatial distributions of targets by capturing motion patterns and structural variations over time. Temporal modeling and motion comprehension enhance robustness against occlusions and interferences, significantly elevating performance [6]. Moreover, motion understanding can support trajectory prediction, aiding downstream tasks like planning and decision-making.

Three primary categories of temporal modeling and motion understanding for multitarget detection and tracking are distinguished: convolution-augmented recurrent networks, attention-based architectures with temporal encoders, and explicit optical-flow estimators. The first class integrates recurrent units within convolution-based detectors or trackers to capture long-range dependencies without enlarging the computation footprint. The second class concatenates temporal encodings to high-dimensional features and employs attention mechanisms to retrieve relevant cues and facilitate multimodal fusion. The third class explicitly computes optical-flow fields to examine intricate motion patterns and allows target-moving applications and adjacent targets to share motion characteristics mutually[18].

8. DATA ASSOCIATION AND IDENTITY PRESERVATION

Data association is a critical task in the Multiple Object Tracking (MOT) area. Given detections arriving over time and a bank of existing tracks, the task is to identify which detections correspond to already tracked objects. The association process is often viewed as an optimization problem [7] , with a common formulation of minimizing a cost function that encodes how likely it is for one detected object to be the same as another one already tracked. A minimalist approach for MOT, tracking-by-detection, breaks the problem into two steps [8]: first, performing object detection over time, and then associating

detections that belong to the same object to form tracks. Hence, new detections are assumed to belong to known objects instead of creating new object instances. The data association stage can often be addressed as a tracking problem that relies on learned models to further aid the identification of already tracked objects [9].

In the MOT context, a tracked object is generally assumed to be an instance of some specific object category, such as pedestrians or vehicles. Yet, during the tracking process, objects often become occluded or change their appearance due to sensor variations. To overcome such challenges, two categories of information can help the data-association task. First, it can exploit motion information by predicting the position to locate the current detection based on the previously associated detection and the associated motion model. Second, it can exploit appearance information by predicting the degree to which a detection from the current frame matches the detected object framed in the previously selected detection [19].

9. DOMAIN ADAPTATION, GENERALIZATION, AND ROBUSTNESS

Deep learning methods frequently rely on large datasets for training. Consequently, trained models may not generalize well when deployments involve different conditions, cameras, scenes, or object categories. Domain adaptation can mitigate problems arising from such domain shifts [10].

Multisensor operation is crucial for applications requiring simultaneous tracking of many people in large areas. Trained detectors may not generalize well across sensors, leading to degraded performance in multisensor scenarios and considerable underutilization of sensor data. Attention mechanisms can alleviate this issue [11]. Detection models normally make

Severe over-fitting may occur in applications with only a few available training images. Artificially generated data helps.

10. REAL-TIME CONSIDERATIONS AND DEPLOYMENT

Real-time operation is often crucial for target detection and tracking. For example, autonomous driving systems must ensure that responses to detected vehicles, cyclists, and pedestrians occur before the tracked object changes lanes, turns, or otherwise becomes occluded [4]. Real-time surveillance systems monitor entrances and exits of restricted zones, which may require complementary tracking of obscured targets. Camera devices, such as drones, body-mounted and handheld infrared cameras, may impose extreme tolerances on algorithmic frameworks, prioritizing low-latency, high-throughput operation and smaller model sizes. Envisioning such constraints illuminates the necessity of extensive and innovative work, which coalesces into four review-level sections[20].

Latency and throughput underpin scheduling throughout methodology delineation: Section 6 surveys the influence of components on detection-and-tracking cycles and consequently models resource requirements. Model-compression techniques, obligatory even with general-size methods, are introduced in section 7, as are other compressor classes. The author and community readership tracker rely on a significant, but not standard, number of released models [21].

Lowering the size, latency, or other aspects of real-time detection-and-tracking systems mandates careful consideration of data-maintenance strategies. When available, pre-compressed datasets stored in raw, high-rate formats fully regain quality prior to later stages. Alternative data-specification regimes—controlling camera speed, intensity, and reshaping through integer-division—are added upon release, supporting further area dispersion when combined with an app. [22].

11. ETHICAL AND SOCIETAL IMPLICATIONS

Group-based deep learning techniques have become increasingly popular due to technological advances and societal changes. However, the widespread use of such technologies also gives rise to ethical and negative societal implications. The emergence of large-scale and digital internet-focused video-acquisition systems leads to the social acceptance of wider verification, detection, and monitoring systems. Nevertheless, the free and unrestricted development of many deep-learning solutions provides opportunities to violate privacy and further contribute to totalitarianism. `Smart cities` and `smart factories` equalize the power of global giants and open the door for totalitarianism under the auspices of terrorism. Different individuals or groups in the world may even utilize the same deep-learning architecture for opposite goals and purposes. Generally, the emergence of deep-learning-based, multitarget detection and tracking technologies necessitates elaborate riskMT mitigation and the

establishment of global standards. Many inherent risks, including personal safety and working conditions, still exist in workplaces. The Internet-based live video broadcast is rapidly converting into a new version of reality show. Various appetites and temptations of manipulation also complete the detection and tracking system of this target. Systems of fairness, neutrality, and balance should be pursued to promote the cooperative balance of multiple and multi-object technologies [23].

12. SYNTHESIS OF METHODOLOGIES

Detection and tracking of multiple objects have advanced significantly, with extensive research and the emergence of diverse deep learning approaches. Rather than presenting an exhaustive survey, this section synthesizes relevant methodologies, offering a perspective on architectural choices and data dependencies. Detection and tracking components from the reviewed work are integrated, enabling systematic juxtaposition of datasets, benchmarks, evaluation metrics, and performance under uniform conditions. Connections among architectures, data, and evaluation frameworks underscore commonalities across methodological paradigms. Such a consolidated view remains valuable for prospective researchers navigating the multifaceted literature on multitarget detection and tracking [24].

13. FUTURE DIRECTIONS

Despite the progress in multitarget detection and tracking with deep learning, challenges remain [1]. Efficient, effective tracking of small objects, handling changing object counts, maintaining identities, managing occlusions, and differentiating similar appearances still require further attention. Continuous growth in data-acquisition platforms for automated driving systems and in the size and diversity of publicly available datasets contribute significantly to the advancement of research in this domain [12]. Future developments are expected to refine and expand object detection and tracking capabilities, thereby improving reliability across real-world scenarios[25].

The establishment of credible standards for the evaluation of data-driven methodologies and the reproducibility of experimental results is essential to the common advancement of detection, tracking, and joint detection-and-tracking methodologies. Current solutions remain largely confined to either single-box or bounding-box-containing tasks. The acquisition of training data requires significant investment of time and resources, and even the most widely used datasets for evaluation remain far from exhaustive. Reproducibility, compatibility among methods, and cross-method evaluation are further hampered by variations in common metrics and experimental conditions. Data- and resource-efficient approaches also constitute a promising research avenue for expanding the access of academia and industry alike to detection, tracking, and data association[24].

14. CONCLUSION

A total of 215 manuscripts were reviewed to document the remarkable progress made by deep learning. Since 2015, a substantial volume of literature on multitarget detection and tracking has emerged, with dedicated surveys each year.

Multitarget detection and tracking are inherently difficult tasks that play significant roles in several applications, such as video surveillance, smart cities, autonomous driving, and augmented reality. For an exhaustive understanding of multitarget detection and tracking and their principles, the associated insightful papers are recommended.

REFERENCES

[1] B. Mirzaei, H. Nezamabadi-Pour, A. Raoof, and R. Derakhshani, “Small object detection and tracking: A comprehensive review,” *Sensors*, vol. 23, no. 15, p. 6887, 2023. https://doi.org/10.3390/s23156887

[2] Z. Soleimanitaleb and M. A. Keyvanrad, “Single object tracking: A survey of methods, datasets, and evaluation metrics,” *arXiv preprint*, 2022.

[3] X. Wu, D. Sahoo, and S. C. H. Hoi, “Recent advances in deep learning for object detection,” *Neurocomputing*, vol. 396, pp. 39–64, 2020. https://doi.org/10.1016/j.neucom.2018.12.090

[4] V. M. Scarrica, C. Panariello, A. Ferone, and A. Staiano, “A hybrid approach to real-time multi-object tracking,” *Applied Sciences*, vol. 13, no. 4, 2023. https://doi.org/10.3390/app13042192

[5] L. Leal-Taixé *et al.*, "Tracking the trackers: An analysis of the state of the art in multiple object tracking," *arXiv preprint arXiv:1704.02781*, 2017.

[6] F. Meng *et al.*, "Spatial–semantic and temporal attention mechanism-based online multi-object tracking," *IEEE Access*, vol. 8, pp. 103671–103681, 2020. https://doi.org/10.1109/ACCESS.2020.2999397

[7] K. Yoon, D. Y. Kim, Y. C. Yoon, and M. Jeon, "Data association for multi-object tracking via deep neural networks," *Sensors*, vol. 19, no. 3, p. 559, 2019. https://doi.org/10.3390/s19030559

[8] M. J. Gómez-Silva, "Deep multi-shot network for modelling appearance similarity in multi-person tracking applications," Ph.D. dissertation, Univ. of Málaga, 2020.

[9] S. J. Sun *et al.*, "Deep affinity network for multiple object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 104–119, 2021. https://doi.org/10.1109/TPAMI.2019.2925760

[10] M. Khodabandeh *et al.*, "A robust learning approach to domain adaptive object detection," *ICCV Workshops*, 2019. https://doi.org/10.1109/ICCVW.2019.00126

[11] L. T. Nguyen-Meidine *et al.*, "Incremental multi-target domain adaptation for object detection with efficient domain transfer," *Pattern Recognition*, vol. 116, 2021. https://doi.org/10.1016/j.patcog.2021.107929

[12] M. Bashar *et al.*, "Multiple object tracking in recent times: A literature review," *IEEE Access*, vol. 10, pp. 101371–101394, 2022. https://doi.org/10.1109/ACCESS.2022.3206690

[13] Z. Zou *et al.*, "Object detection in 20 years: A survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257–276, 2023. https://doi.org/10.1109/JPROC.2023.3242089

[14] Z. Q. Zhao *et al.*, "Object detection with deep learning: A review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, 2019. https://doi.org/10.1109/TNNLS.2018.2876865

[15] R. Kaur and S. Singh, "A comprehensive review of object detection with deep learning," Digital Signal Processing, vol. 132, p. 103812, 2023. https://doi.org/10.1016/j.dsp.2022.103812

[16] X. Zou, "A review of object detection techniques," in *Proc. ICSGEA*, 2019, pp. 251–254. https://doi.org/10.1109/ICSGEA.2019.00060

[17] K. Li and L. Cao, "A review of object detection techniques," in *Proc. ICECTT*, 2020, pp. 385–390. https://doi.org/10.1109/ICECTT48285.2020.9115734

[18] X. Wu, D. Sahoo, and S. C. H. Hoi, "Recent advances in deep learning for object detection," *Neurocomputing*, vol. 396, pp. 39–64, 2020. https://doi.org/10.1016/j.neucom.2018.12.090

[19] B. Mirzaei *et al.*, "Small object detection and tracking: A comprehensive review," *Sensors*, vol. 23, no. 15, p. 6887, 2023. https://doi.org/10.3390/s23156887

[20] T. Shehzadi *et al.*, "Object detection with transformers: A review," *Sensors*, vol. 25, no. 19, p. 6025, 2025. https://doi.org/10.3390/s25196025

[21] Y. Liu *et al.*, "A survey and performance evaluation of deep learning methods for small object detection," *Expert Systems with Applications*, vol. 172, 2021. https://doi.org/10.1016/j.eswa.2021.114602

[22] K. U. Sharma and N. V. Thakur, "A review and an approach for object detection in images," *Int. J. Computational Vision and Robotics*, vol. 7, no. 1–2, pp. 196–237, 2017. https://doi.org/10.1504/IJCVR.2017.083276

[23] A. K. Shetty *et al.*, "A review: Object detection models," in *Proc. I2CT*, 2021. https://doi.org/10.1109/I2CT51068.2021.9418034

[24] D. K. Prasad, "Survey of the problem of object detection in real images," *International Journal of Image Processing*, vol. 6, no. 6, p. 441, 2012.

[25] A. Borji *et al.*, "Salient object detection: A survey," *Computational Visual Media*, vol. 5, no. 2, pp. 117–150, 2019. https://doi.org/10.1007/s41095-019-0139-9