







Facing challenges: A survey of object tracking [☆]

Wenqi Zhang ^{a, }, Xinqiang Li ^{b, }, Xingyu Liu ^{c, }, Shiteng Lu ^{a, }, Huanling Tang ^{a,d,e, },*

^a School of Computer Science and Technology, Shandong Technology and Business University, Yantai, Shandong, 264005, China

^b Shandong Luruan Digital Technology Co., Ltd. Smart Energy Branch, Jinan, Shandong, 250001, China

^c Monash University Joint Graduate School (Suzhou), Southeast University, Suzhou, Jiangsu, 215123, China

^d Co-Innovation Center of Shandong Colleges and Universities: Future Intelligent Computing, Yantai, Shandong, 264005, China

^e Key Laboratory of Intelligent Information Processing in Universities of Shandong, Shandong Technology and Business University, Yantai, Shandong, 264005, China

ARTICLE INFO

Keywords:

Computer vision
Object tracking
Deep learning
Convolutional neural network
Transformer

ABSTRACT

Object tracking, regarded as one of the most fundamental and challenging problems in computer vision, has attracted considerable attention in recent years. Researchers have conducted extensive studies on object tracking. However, research focused on challenges is rare. This paper concentrates on the challenges of object tracking in different periods, analyzing the reasons behind these challenges. In this paper, we have also consulted related work on object tracking from multiple aspects, including solutions based on challenges, popular application directions, and future prospects. It is expected that this paper will offer valuable references for researchers, and promote the innovation and advancement of object tracking technology in various fields.

1. Introduction

Object tracking is an important and popular research direction in computer vision [1]. It is a technology that continuously tracks the moving and changing object in the scene by analyzing image sequences. It involves image processing [2], object detection [3], and other disciplines [4], and has a wide range of applications in virtual reality [5], autonomous driving [6], face recognition [7] and other fields [8]. Based on the varying number of objects being tracked (one or multiple), object tracking can be further divided into two significant branches: Single Object Tracking (SOT) and Multiple Object Tracking (MOT), as shown in Fig. 1.

Object tracking technology has made remarkable progress, but it still faces many challenges. From 2010 to 2015, object tracking encountered the challenge of feature representation [9]. Accurate feature representation is fundamental to object tracking, with traditional object tracking algorithms predominantly relying on feature extraction and representation. To address this challenge, researchers have proposed numerous object tracking algorithms. For instance, the tracking method based on the target motion model can describe accurately the actual movement law of the object [10]. During this period, research on object tracking was primarily focused on the algorithm itself. However, this approach

had its limitations, especially when dealing with complex environments or large-scale requirements.

Between 2015 and 2020, with the rapid development of hardware devices and the widespread application of deep learning technology, many researchers further applied deep learning technology to object tracking [11][12]. For instance, the Recurrently Target-attending Tracking method captures long-distance information by utilizing a recurrent neural network with multiple directions to improve tracking performance [13]. The object tracking algorithms based on deep learning can effectively improve the robustness of object tracking, especially in complex scenes [4][14].

In the 2020s, with the increasing demand for data fusion and large-scale object tracking, not only have the accuracy requirements for object tracking feature representation been improved, but the real-time inference requirements for large-scale object tracking have also increased. Recent studies [15][16] have demonstrated that deep learning-based object tracking methods can address the real-time inference challenges of object tracking. For instance, a 2024 study [17] proposed the TransTRDT model, which improves inference speed by incorporating dynamic template. Despite the remarkable achievements in the field of object tracking, there are still environmental challenges that remain

[☆] This work was supported by the National Natural Science Foundation of China (No.62341605, No.62372077).

* Corresponding author at: School of Computer Science and Technology, Shandong Technology and Business University, Yantai, Shandong, 264005, China.

E-mail addresses: 2023420181@sdtbu.edu.cn (W. Zhang), iiangiang@163.com (X. Li), liuxingyu_025@seu.edu.cn (X. Liu), 2023420196@sdtbu.edu.cn (S. Lu), thL01@163.com (H. Tang).



Fig. 1. Display of different object tracking tasks. The blue dashed box signifies the SOT scenario, which involves only one tracked object. Conversely, the red dashed box denotes the MOT scenario, encompassing multiple tracking objects.

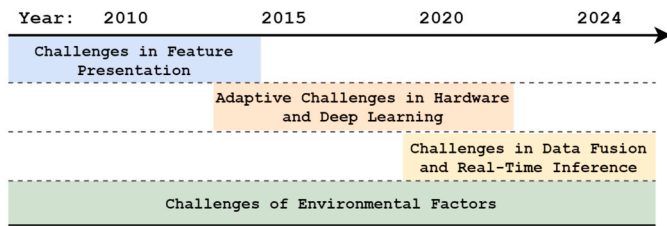


Fig. 2. The challenges encountered by object tracking vary across different periods. Object tracking faced the challenge of feature representation during the period from 2010 to 2015. From 2015 to 2020, it was challenged by the adaptation to hardware and deep learning frameworks. From 2020 to the present, it has faced the challenges of data fusion and real-time inference. The challenge posed by environmental factors is a persistent one that object tracking has to encounter throughout its entire development process.

unsolved. For example, changes in object appearance, nonlinear deformation can affect the accuracy and stability of object tracking.

Reviewing object tracking based on challenges is of great significance, yet currently, research in this area remains insufficient. Therefore, this paper focuses on the challenges faced by object tracking from 2010 to the present, and analyzes these challenges across different periods, as shown in Fig. 2. For this paper, the main contributions are as follows:

- (1) Analyzing various challenges faced by object tracking in different periods and their underlying reasons.
- (2) Thoroughly summarizing and contrasting the solutions to diverse challenges.
- (3) Formulating outlooks concerning the challenges that need to be addressed and potential research directions.

2. Background

2.1. Conventional object tracking algorithms

In 2010s, object tracking algorithms primarily concentrated on the basic tracking problem. Over time, this field has evolved, transitioning from simple to complex models and from linear to nonlinear approaches. During this period, object tracking techniques mainly relied on fundamental image processing and pattern recognition methodologies. On the

one hand, the application of Wiener filters to prediction and signal separation represented the initial use of linear filtering techniques [18]. The essence of the Wiener Filter is to minimize the mean square value of the difference between the desired response and the actual output of the filter. On the other hand, the application and principle of the conventional Kalman Filter in object tracking are primarily rooted in its capability to handle nonlinear systems and its efficient computational performance. This algorithm is utilized to mitigate the challenges associated with nonlinear problems in object tracking [19]. Specifically, the conventional Kalman Filter is an algorithm that uses a linear system state equation to estimate the system state based on the input and output observation data.

Both Optical Flow and Feature Flow constitute significant motion description methodologies in the field of computer vision. Optical Flow elucidates the temporal motion state of pixels within an image, encompassing the motion vector of each pixel, which details its speed and direction [20]. Conversely, Feature flow refers to the temporal movement state of feature points in an image, specifically emphasizing the obvious features of these points [21].

With the advancement of computer science and engineering techniques, object tracking algorithms have begun to shift to more complex systems. For instance, the Interacting Multiple Model (IMM) [22] was proposed to address the challenges of filtering linear systems, marking a significant advancement compared to traditional filtering methodologies. Meanwhile, algorithms such as Probabilistic Data Association (PDA) and Joint Probabilistic Data Association (JPDA) have been developed to facilitate robustness of object tracking in highly cluttered environments [23].

2.2. Hardware advancement

2.2.1. Computing power platform

There exist two primary types of computing power platforms: Central Processing Unit (CPU) and Graphics Processing Unit (GPU). The transition from CPU to GPU is a complex and multifaceted process, involving the evolution of hardware architectures and shifts in computing demands. The approach of boosting processor performance by increasing the number of CPU transistors has reached its physical limits, due to the challenges posed by Moore's Law and Dennard Scaling [24]. Consequently, this has spurred the investigation of novel architectures, within

Table 1

Performance comparison with state-of-the-art object tracking methods based on deep learning across the different datasets.

Type	Method	TrackingNet			LaSOT		
		<i>AUC</i>	<i>P_{norm}</i>	<i>P</i>	<i>AUC</i>	<i>P_{norm}</i>	<i>P</i>
SOT	TransT (Chen et al. 2021) [46]	81.4	86.7	80.3	64.9	73.8	69.0
	STARK (Yan et al. 2021) [47]	82.0	86.9	-	67.1	77.0	-
	SwinTrack (Lin et al. 2022) [48]	82.5	87.0	80.4	69.6	78.6	74.1
	ToMP (Mayer et al. 2022) [49]	81.2	86.2	78.6	67.6	78.0	72.2
	MixFormer (Cui et al. 2022) [50]	82.6	87.7	81.2	67.9	77.3	73.9
	TATrack (He et al. 2023) [16]	85.0	89.3	84.5	71.1	79.1	76.1
	TransTRDT (Sun et al. 2024) [17]	85.6	-	84.9	72.1	-	75.4
	MixFormerV2 (Cui et al. 2024) [51]	83.4	88.2	81.6	70.6	80.8	76.2
Type	Method	MOT17			DanceTrack		
		<i>HOTA</i>	<i>MOTA</i>	<i>IDF1</i>	<i>HOTA</i>	<i>MOTA</i>	<i>IDF1</i>
MOT	TransTrack (Sun et al. 2020) [52]	54.1	74.5	63.9	45.5	88.4	45.2
	GTR (Zhou et al. 2022) [53]	59.1	75.3	71.5	48.0	84.7	50.3
	Bytetrack (Zhang et al. 2022) [54]	63.1	80.3	77.3	47.7	89.6	53.9
	MeMOTR (Gao et al. 2023) [15]	58.8	72.8	71.5	68.5	89.9	71.2
	C-BIoU (Yang et al. 2023) [55]	64.1	81.1	79.7	60.6	91.6	61.6
	OC-SORT (Cao et al. 2023) [56]	63.2	78.0	77.5	55.1	92.0	54.6
	SMILEtrack (Wang et al. 2024) [57]	65.3	81.1	80.5	-	-	-
	Motiontrack (Xiao et al. 2024) [58]	-	-	-	58.2	91.3	58.6

which GPU has emerged as a pivotal research area due to its high level of parallelism and relatively low power consumption [25][26].

With the advancement of technology, GPUs have evolved from being mere graphics processing units to multi-core processors capable of supporting general computing. This transformation has not only retained their crucial role in graphics rendering, but has also unlocked considerable potential in scientific computing, deep learning, and numerous other fields [27].

2.2.2. Sensing device

From the camera perspective, the monocular camera and the multi-lens camera represent two prevalent types of cameras utilized within the domain of object tracking. They exhibit notable differences in functionality, application scenarios, and performance. A monocular camera, characterized by a single lens, acquires scene information by capturing images from a single viewpoint. In contrast, a multi-lens camera, comprising multiple lenses, possesses the capability to capture images from varied viewpoints [28]. Consequently, it is able to estimate the depth and position of objects more precisely. Given that multi-lens cameras provide richer scene information and exhibit superior positioning accuracy, they perform outstandingly in applications such as video surveillance [29] and eye tracking [30]. Compared with traditional RGB cameras, depth cameras offer depth information of the scene, thereby facilitating a more effective resolution of object tracking challenges [31]. Meanwhile, by acquiring 3D information of the scene, depth cameras provide additional spatial dimension data for object tracking. This capability is of importance in mitigating issues such as occlusion and illumination variations within intricate environments [32].

Additionally, commonly utilized sensing devices also include LiDAR, a technology that employs lasers as light sources for precise distance measurement. By transmitting a laser beam and receiving the reflected signal, it determines the range, velocity, and position of the object [33]. LiDAR technology is widely employed in numerous fields due to its excellent accuracy, high resolution, and extensive long-range detection capabilities [34].

2.3. Deep learning

The feature extraction capability of deep learning stands as one of the pivotal factors contributing to its widespread success. Through continuous research and innovation, deep learning has made significant strides in numerous domains, including image recognition and natural

language processing [35]. With the rapid advancement of deep learning technology, its potent feature extraction ability is now extensively utilized in the field of object tracking. Deep learning models, such as convolutional neural networks (CNN) and Transformers [36][37], are utilized to extract comprehensive and detailed feature information from objects, thereby enhancing the overall performance of object tracking systems [38]. In addition, there are some classical deep learning frameworks related to object tracking. For instance, R-CNN [3][39] initially utilizes a selective search algorithm to generate candidate regions and extract features. It then employs bounding box regression to adjust the position of these regions. Subsequently, it filters out overlapping candidate regions to obtain the final detection results. On the other hand, YOLO [40], which is based on a modified architecture inspired by U-Net [2], leverages CNN to extract image features and directly predict the bounding box coordinates and class probability of the object. This approach enhances the accuracy and robustness of detection. Furthermore, YOLO also can be adapted to various other computer vision tasks [41][42].

Deep learning significantly improves the accuracy and robustness of object tracking by automatically extracting high-dimensional features from image or video sequences [35]. It also effectively addresses long-term tracking challenges arising from object occlusion and limited field of view [43]. The performance comparisons of different object tracking methods based on deep learning are presented in Table 1. Additionally, by enhancing the quality of images during the object tracking process, techniques such as image super-resolution [44] and image inpainting [45] can effectively alleviate challenges like object occlusion or background clutter, to some extent.

3. Challenges and reasons

This chapter comprehensively analyzes the latest researches on object tracking and elaborates in detail on the challenges faced in this field, along with the underlying reasons. Table 2 presents an overview of these challenges and the problems they entail.

3.1. Challenge 1: feature representation

The feature representation of object tracking poses a significant challenge, as it necessitates the precise extraction and representation of an object's appearance features from video sequences to ensure accurate tracking. During the 2010s, advancements in hardware technology and

Table 2
Challenges and Problems in Object Tracking.

Challenges	Problems
Challenge 1: Feature Representation	1. Failure to accurately extract and represent features
Challenge 2: Adaptation of Hardware and Deep Learning	1. Tracking drift phenomenon 2. Image void or artifact issue 3. Ineffectively copes with appearance alterations 4. Insufficient generalization ability 5. Complex Data association
Challenge 3: Data fusion and Real-Time Inference	1. Fails to integrate large-scale data effectively 2. Unable to achieve effective real-time prediction
Challenge 4: Environmental factors	1. Occlusion and deformation 2. Fast-moving object 3. Illumination change

computing power lagged behind, causing object tracking techniques of that era to heavily rely on fundamental motion models for predicting object locations. Consequently, these models typically lack the capability to accurately represent features, resulting in their inability to track objects precisely [9][59].

3.2. Challenge 2: adaptation of hardware and deep learning

3.2.1. Challenge 2a: adaptation of hardware

From the processor's perspective, with the rapid technological advancements, GPUs have emerged as superior in parallel processing capabilities compared to CPUs, gradually dominating the realm of artificial intelligence computing. However, despite significant enhancements in GPU performance, differences between them and CPUs still persist [60]. Therefore, to fully harness the capabilities of the GPU, there is a pressing need for the development of more efficient parallel programming models and tools, accompanied by optimization strategies for specific application areas [61].

From the camera's perspective, a primary limitation of monocular cameras lies in their inability to directly provide depth information. Moreover, when confronted with rapidly moving objects, monocular cameras may suffer from tracking drift [62]. In contrast, multi-lens cameras, which capture images from multiple viewpoints concurrently, can gather more extensive scene details. However, these equipments are also affected by factors such as lens resolution, field of view, and image synchronization [63]. Conversely, traditional cameras are unable to record the distance between the object and the camera. By measuring the depth of field within the captured space, depth cameras can capture the distance of each point from the camera. However, depth cameras also face challenges. Environmental factors can adversely affect the pose estimation accuracy of depth cameras [64]. Additionally, depth cameras also encounter other challenges such as holes and artifacts in depth images, and flickering in depth videos [65].

From the perspective of LiDAR, LiDAR point cloud tracking has consistently sparked the curiosity of numerous researchers. The core of point cloud object tracking lies in the extraction of effective features from point cloud data, thereby enabling precise tracking [66][67]. The irregular and sparse nature of point cloud data poses significant challenges to object tracking. Consequently, due to its lack of continuity and texture information similar to images, traditional image-based object tracking methods are not readily applicable to point cloud data [67].

3.2.2. Challenge 2b: adaptation of deep learning

The challenges faced in adapting object tracking and deep learning are primarily evident in the following three aspects:

(1) Handling Object Appearance Variations: Throughout the tracking process, the object may undergo deformation, rapid motion, illumination changes, and long-term occlusion, thereby placing greater demands on the robustness of deep learning-based object tracking algorithms [68]. Although deep learning techniques can improve the adaptability

to these variations by extracting deep features of the object, effectively managing complex changes remains a challenge.

(2) Model Adaptation and Generalization Ability: Deep learning-based object tracking algorithms must adapt to changes in the object and maintain strong performance in different environments. However, existing deep learning object tracking algorithms often rely heavily on a large volume of labeled data for training, which limits their generalization ability when sufficient labeled data is unavailable [69].

(3) Data Association and MOT: In MOT scenarios, one of the significant challenges for deep learning object tracking is accurately associating the detected object with its previous trajectory [70]. This challenge represents a prevalent and active research direction in the field of object tracking.

3.3. Challenge 3: data fusion and real-time inference

3.3.1. Challenge 3a: data fusion

With the widespread adoption and application of surveillance and mobile terminal equipment, a substantial volume of video data emerges in our daily lives, providing indispensable data and technical support for object tracking [68]. Nevertheless, the analysis, processing, and fusion of these data encounter considerable challenges.

On the one hand, object tracking techniques need to handle large-scale data. For instance, the presence of background noise and other factors often undermines the reliability of moving object detection in extensive video sequences. When these detected objects serve as candidate regions for tracking, it frequently leads to problems such as object tracking loss and drift [71][72].

On the other hand, the diversity and complexity of data are constantly increasing. As sensing environment and objects become more complex and variable, object tracking must constantly adapt to uncertain, nonlinear, and multimodal complex systems [73].

3.3.2. Challenge 3b: real-time inference

In scenarios such as video surveillance and autonomous driving, object tracking necessitates the real-time prediction of an object's position and motion state within video or image sequences. Nevertheless, deep learning-based object tracking algorithms are often computationally intensive, posing difficulties in achieving real-time inference [69][74][75]. Hence, there remains a pressing need for further research into methods that can enhance real-time inference capabilities while maintaining high accuracy.

3.4. Challenge 4: environmental factors

Object tracking technology encounters numerous challenges when dealing with environmental changes. These challenges are primarily due to the interaction between the object and its background, as well as variations within the object itself, as depicted in Fig. 3.

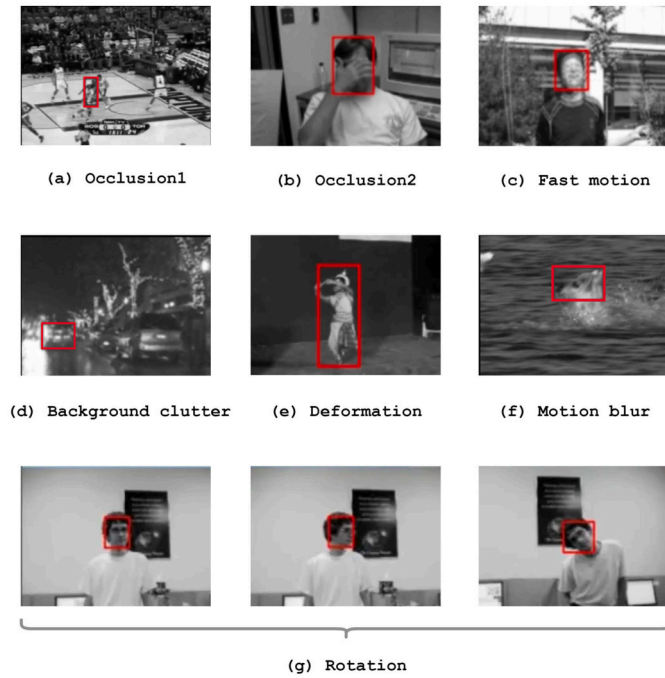


Fig. 3. The manifestations of challenges posed by environmental factors. (a)(b) indicate the object is occluded. (c) reveals the rapid movement of the object. (d) represents the object in complex backgrounds. (e) shows the deformation of the object. (f) indicates the object motion blur. (g) demonstrates the rotation of the object.

3.4.1. Challenge 4a: occlusion and deformation

Occlusion and deformation represent the most prevalent environmental challenges in object tracking. When a object is occluded by other objects during tracking, it becomes impossible to fully observe the target [1][43]. Additionally, Deformation implies that the object's shape and scale may undergo changes within the video sequence, subsequently affecting the tracking accuracy [4][10].

3.4.2. Challenge 4b: fast-moving object

When the object is moving at a high speed, its position varies significantly between adjacent frames, thereby increasing the difficulty of predicting its position in the current frame [10][73].

3.4.3. Challenge 4c: illumination change

The variation in illumination renders the feature extraction of object tracking challenging, aggravates the occlusion problem and rendering visual features more sensitive [4]. Additionally, in low-light environments, pixel values undergo alterations, leading to the weakening of the object's texture and color features [76].

4. Solutions for feature representation

Regarding the challenge of feature representation in object tracking, this chapter summarizes the related research based on traditional object tracking algorithms and simple mathematical models. Although these algorithms may appear somewhat outdated in the current technological landscape.

4.1. Particle filter

Particle filter is a pivotal technique in object tracking, especially for addressing nonlinear and non-Gaussian problems. This algorithm estimates the object's state by approximating the posterior probability density through a set of random samples [77][78]. Within the realm of object tracking, the Particle filter faces the challenge of feature representation. To address this, researchers have proposed various improvement

methods. For instance, self-refactoring Particle filter addresses complex and variable tracking environments by splitting and merging trackers, thereby improves the tracking efficiency and accuracy [79]. Another approach, the multi-moving objects tracking algorithm based on feature points, identifies multiple moving objects using an adaptive background modeling algorithm and integrates the Scale-Invariant Feature Transform (SIFT) algorithm for motion tracking [59].

4.2. Kernel correlation filtering

Correlation filtering algorithms, particularly the Kernel Correlation Filter (KCF) algorithm, have attracted significant attention due to their remarkable efficiency. The KCF algorithm leverages the Fast Fourier Transform to accelerate computations, thereby enabling the algorithm to significantly boost its capacity to process high-dimensional features while maintaining real-time performance [80]. The improved KCF algorithm [81] builds the Scale Pyramid for the current object and subsequently acquires the object's current scale information through the utilization of a Multi-Feature Map Kernelized Correlation Filter.

4.3. Kalman filter

The Kalman Filter is a widely employed algorithm in tracking and estimation, favored for its simplicity, optimality, processability, and robustness. Object tracking algorithms based on KF also alleviate the challenges of feature representation in the field of object tracking. For instance, adaptive correlation filters and Kalman filtering system [82], effectively extract key features by conducting multi-object detection within reduced frame areas. The modified Kalman filter [83] robustly establishes a mapping between unordered detected objects and Kalman estimates, further enhancing the feature representation capabilities of object tracking.

4.4. Algorithms based on optical flow or feature flow

The Feature Flow estimation module within IFF-Net [21] possesses the capability to directly produce feature flow, which represents the movement of features within the network. This module shares features with the detection branch, enabling it to efficiently capture object motion information. The object tracking algorithm that integrates particle filter and optical flow methodologies utilizes the optical flow technique to ascertain the motion vector of the object in the initial frame. Subsequently, it predicts the potential location of the object and disseminates particles in its vicinity [84]. Throughout the tracking process, parameters, including the particle count, are dynamically adjusted as necessary to enhance the tracking performance.

5. Solutions for adaptation of hardware and deep learning

Regarding the adaptable challenges posed by object tracking in the context of hardware and deep learning, this chapter delves into the solutions addressing both challenges, as presented in Table 3. Fig. 4 illustrates the processing pipeline of a deep learning-based object tracking system. Fig. 5 depicts the various types of deep learning models.

5.1. Solutions for adaptation of hardware

5.1.1. Computing power platform

With the technological advancements, GPU technology has been extensively utilized, significantly improving the performance and real-time capability of object tracking systems. The implementation of particle filters on Nvidia GPUs utilizing half-precision floating-point numbers has demonstrated substantial performance improvements [89]. By harnessing the combined computing powers of CPUs and GPUs, CPU-GPU integration technology effectively improves the execution speed and efficiency of object tracking algorithms [85]. Furthermore, the evolution

Table 3

Solutions for adaptation of hardware and deep learning.

Model / References	Advantage	Disadvantage
(Florian et al. 2015) [85] (Yu et al. 2017) [86] (Liu et al. 2018) [31] (Ke et al. 2019) [64] (Yan et al. 2020) [62] (Liu et al. 2021) [87] (Sun et al. 2021) [88]	fully leverage GPU acceleration technology better object positioning accuracy efficiently track 3D positions and handle occlusion and illumination changes alleviate the influence of camera pose and improve the quality of 3D model improve the accuracy and recognition rate of object tracking resolve the problem of scene rotation, similar object interference and drift effectively address the tracking drift problem caused by environment interference and occlusion implement Particle filter algorithm within the GPU framework	— affected by environment great computational complexity prolonged algorithm runtime the correlation between modules is poor incapable of resolving long-term occlusion issues incapable of handling multi-object detection
(Gabin et al. 2023) [89] RTT (Cui et al. 2016) [13] DSiam (Guo et al. 2017) [90] (Li et al. 2018) [91] (Wu et al. 2018) [92] (Miao et al. 2018) [93] (Li et al. 2019) [94] (Jia et al. 2019) [95] (Li et al. 2019) [96] (Zhang et al. 2019) [97] TransTrack (Sun et al. 2020) [52] UniTrack (Wang et al. 2021) [98] UTT (Ma et al. 2022) [99]	suppresses a portion of the clutter background information adapt to temporal variations in foreground and background reduce the complexity and enhance efficiency better speed and performance fusion of multi-layer features better accuracy and efficiency efficiently handle severe occlusion and interference from similar objects effectively handle occlusion and blurring and increase accuracy effectively address drastic appearance changes and mitigate model complexity achieve joint detection and tracking, improve the performance and speed simultaneously complete SOT and MOT simultaneously achieve better completion of SOT and MOT	— computationally costly the accuracy is relatively low poor accuracy subjected to occlusion complex model architectures large model parameters low algorithm speed time-consuming computation poor accuracy excessive model complexity incapable of handling complex environments demand a large number of training models

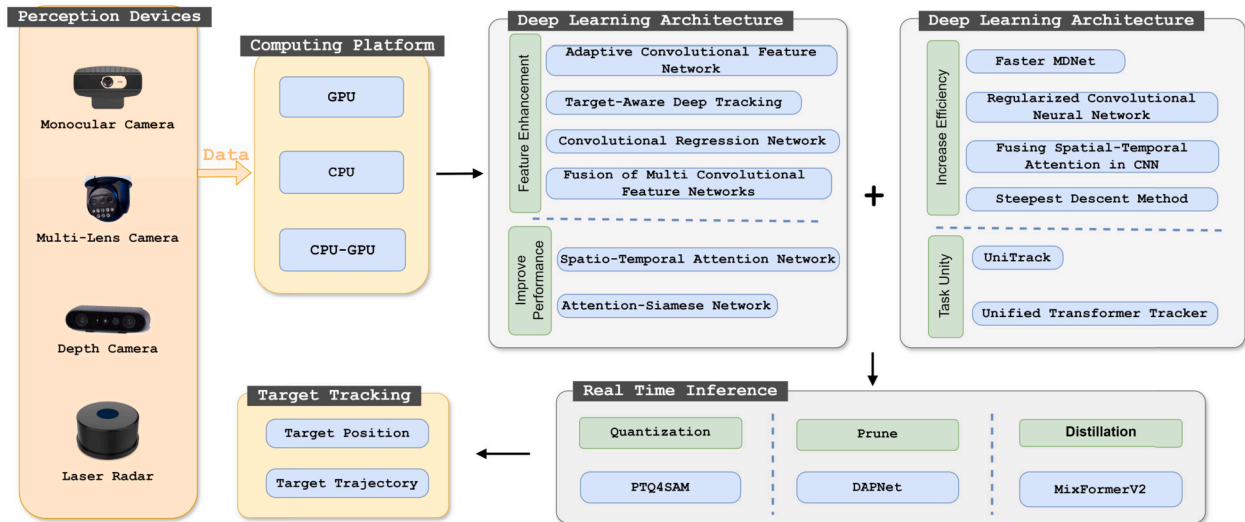


Fig. 4. The solution process of the object tracking method based on deep learning. The data originates from various types of perception devices, and object tracking utilizes this data to operate on computing platforms, integrating different deep learning architectures and employing real-time inference to obtain the final object tracking information.

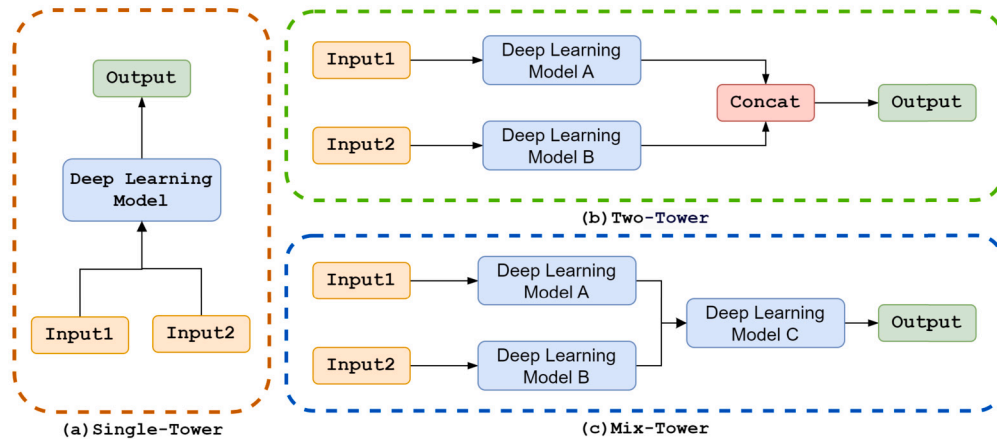


Fig. 5. Different types of deep learning architectures. (a) signifies the single-tower model, characterized by a solitary processing path or structural framework. (b) designates the two-tower model, which encompasses parallel or branching processing pathways. (c) represents the mix-tower model, integrating aspects of both (a) and (b). Models A, B, and C can represent the same or different deep learning models.

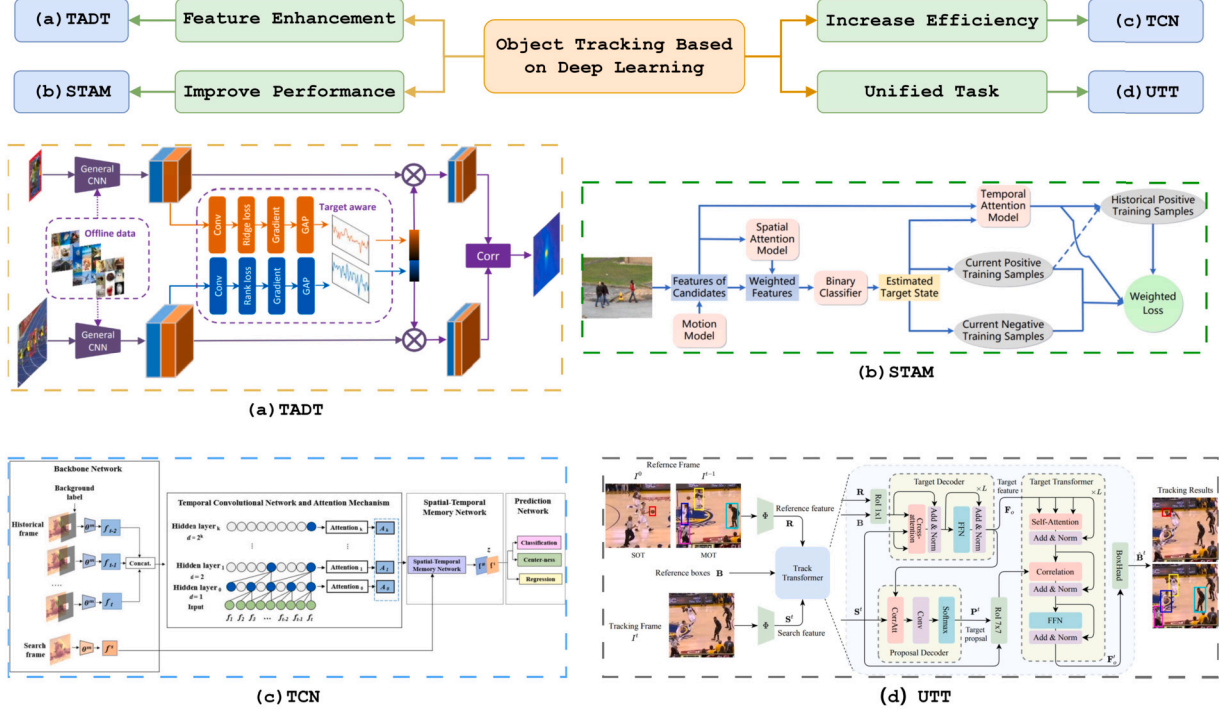


Fig. 6. An overview of various types of object tracking methods based on deep learning frameworks. Each category offers an instance of a classic model. (a) represents the TADT framework [94], (b) represents the STAM framework [100], (c) represents the TCN framework [101], and (d) represents the UTT framework [99].

of GPUs has further propelled the continuous refinement of deep learning technology, facilitating the integration of advanced deep learning frameworks into object tracking systems.

5.1.2. Camera apparatus

Due to the different types of data acquired by monocular and multi-lens camera, researchers have proposed numerous solutions for these distinct data types. For monocular cameras, an enhanced SIFT algorithm [86] has been introduced, which builds upon the principle of feature classification and leverages the properties of corner points to improve the traditional SIFT algorithm, ultimately enhancing the precision of object tracking. For multi-lens cameras, researchers have proposed an object tracking algorithm based on high-resolution Siamese network [62]. This algorithm employs a high-resolution network to extract image features, emphasizing useful features while suppressing unnecessary information. The extracted features are further refined, addressing problems such as tracking drift and positioning inaccuracies in object tracking tasks.

Depth cameras are capable of providing depth information of the scenes, allowing for a more accurate representation of object features. However, they also face several challenges. To address these, researchers have adopted depth camera-based 3D scene reconstruction techniques, fully exploiting the constraints among different objects and planes in the image. A fine-to-coarse camera tracking algorithm has been introduced [64], effectively addressing the problem of camera pose instability [64]. Additionally, a recent study [65] has proposed an image enhancement method leveraging digital image processing techniques, which has alleviated the problems of holes and artifacts in depth maps.

5.1.3. LiDAR

LiDAR point cloud tracking represents a complex and comprehensive research field, serving as a key technology in various fields, including automatic driving, robot navigation, and environmental monitoring. The effective processing of point cloud data forms the bedrock for successful tracking tasks. SimTrack [66] employs an end-to-end trainable model that undertakes joint detection and tracking directly from original point cloud data, thereby simplifying the traditional hand-crafted tracking

approach and improving the accuracy of object tracking. Meanwhile, research on object tracking based on 3D laser point cloud [67] reduces the complexity of original laser point cloud data by introducing point cloud filtering and segmentation algorithms. By integrating these algorithms with machine learning techniques, the 3D laser point cloud data is processed directly, ensuring maximum retention of perceptual information. These approaches significantly boost the stability and accuracy of object tracking in intricate environmental settings.

5.2. Solutions for adaptation of deep learning

The application of deep learning in the domain of object tracking has made significant strides, gradually surpassing traditional methods and emerging as an important research direction within this field. An overview of various deep learning-based object tracking methods is presented in Fig. 6.

5.2.1. Feature enhancement

The integration of object tracking with deep learning through feature enhancement can be highly effective. Below are refined descriptions of various methods:

(1) The object tracking algorithm based on adaptive convolutional features [102][91] conducts principal component analysis and dimensionality reduction on the features extracted from specific layers of CNN. This approach enables the model to learn more diversified feature representations.

(2) Target-Aware Deep Tracking [94][16] improves tracking performance by learning target-aware features that enhance the ability to identify and distinguish the object of interest. This method leverages a combination of fusion regression loss and ranking loss to select the most effective feature representation.

(3) The Convolutional Regression Network [95] algorithm with a fusion redetection mechanism integrates the correlation filter as a CRN layer into the deep neural network. This allows the network to be trained end-to-end, improving feature quality.

(4) The moving object tracking algorithm [96], which leverages multi-layer convolutional feature correlation filtering, integrates the

Table 4

Common datasets for SOT and MOT. N indicates that the number of tracked objects is greater than one. The sizes of all datasets may vary slightly. Given that the VOT dataset is updated annually, this table takes VOT2016 as a representative example.

Tracking Types	Dataset Name	Videos	Object Types	FPS	Size	Objects
SOT	OTB100 [106]	100	-	-	2.54 GB	1
	VOT2016 [107]	60	-	-	1.25 GB	
	UAV123 [108]	123	-	30-96	13.5 GB	
	TrackingNet [109]	30,643	27	-	1.14TB	
	GOT-10k [110]	10,000+	563	-	66 GB	
	LaSOT [111]	1400	70	-	227 GB	
MOT	MOT17 [112]	14	12	30	5.5 GB	N
	RGB-D Object [113]	-	51	-	84 GB	
	NYU Depth V2 [114]	-	1000+	-	428 GB	
	MOT20 [115]	8	5	25	5.0 GB	
	TAO [116]	2,907	833	-	347 GB	
	DanceTrack [117]	100	1	20	16.7 GB	

features extracted from all convolutional layers. This fusion process significantly amplifies the richness and effectiveness of the feature representation.

5.2.2. Increase tracking efficiency

The object tracking algorithm based on deep learning frameworks can effectively improve the efficiency of object tracking. For instance, by integrating multi-domain CNNs with the optical flow method, the speed and accuracy of object tracking are improved [103]. Additionally, regularized CNNs have been proposed to augment the robustness and stability of object tracking [97]. A recent study [100] introduces a CNN-based MOT approach that utilizes single-object trackers, sharing features and employing ROI-Pooling to derive distinct features for each object, thereby boosting computational efficiency. Enhancing efficiency stands as a primary objective within the domain of object tracking, and indeed, across the broader scope of artificial intelligence.

5.2.3. Improve detection performance

The object tracking algorithm based on spatio-temporal informations not only significantly enhances the adaptability of object tracking and deep learning frameworks but also boosts the performance of object tracking [101][88][104]. For instance, ThrAtt-Siam [105] integrates channel attention mechanism and spatial attention mechanism into the object tracking framework. The channel attention mechanism is capable of effectively distinguishing useful features from useless ones, and extracting key features across different channels. The spatial attention mechanism, on the other hand, complements the object feature information within the channel space, thereby enhancing the accuracy of object tracking.

5.2.4. Task unity

Object tracking, as a important area of computer vision research, has divided into two distinct research domains: SOT and MOT. Thus, both SOT and MOT face the challenge of adapting to deep learning frameworks [98][99]. Due to the differences in training datasets and tracking objects between these two tasks, as shown in Table 4, current methodologies struggle to adapt to both SOT and MOT. To tackle this problem and facilitate tracking across diverse scenarios, the Unified Transformer Tracker (UTT) [99] has been introduced. UTT provides feature mapping suggestions within the tracking frame, based on prior positioning information, for both SOT and MOT tasks. By associating the object features with these suggestions, UTT updates the object representation and outputs the object's position. This approach alleviates the challenges associated with adapting object tracking to deep learning frameworks.

6. Solutions for data fusion and real-time inference

With the persistent emergence of large-scale data, the processing and fusion of such extensive data have emerged as pivotal challenges

in the realm of object tracking. Inevitably, the handling of massive data presents considerable obstacles for real-time inference. Consequently, in recent times, a substantial body of research has been dedicated to addressing the challenges of data fusion and real-time inference, as shown in Table 5.

6.1. Solutions for data fusion

Data fusion is categorized into temporal data fusion, spatial data fusion and spatio-temporal data fusion. A detailed description of different types of object tracking data is presented in Fig. 7. The fusion of diverse data types poses one of the primary challenges in the field of object tracking.

6.1.1. Temporal data fusion

MeMOT [121] offers a comprehensive framework for object detection and data association, enabling the linkage of objects across extended periods to accomplish data fusion. The system encompasses three core modules:

- Hypothesis generation, responsible for generating object proposals within the current video frame.
- Memory coding, which extracts the essential information of the tracked object.
- Memory decoding, tasked with resolving object detection and data association challenges.

Within the memory coding module, there are three attention mechanisms: the short-time block, focused on attending to adjacent frames; the long-time block, correlating features across multiple frames; and the hybrid block, which performs temporal data fusion on both short and long-time blocks to produce the final tracking result.

6.1.2. Spatial data fusion

MixFormer [50] exploits the flexibility of attention operations and proposes a Mixed Attention Module (MAM) employed for feature extraction and object data fusion. The inputs to MAM consist of the target template and the search area. The objective of MAM is to extract the distinct data features from these inputs and integrate their interactive information, thereby achieving spatial data fusion. Within MAM, self-attention operations are performed on the tokens within each sequence to capture the object or seek specific information. Meanwhile, cross-attention operations are conducted between the target template and the search area, enhancing the interplay of spatial data across different scales and further improving the system's spatial data fusion capabilities.

Table 5

Solutions for data fusion and real-time inference.

Model / References	Advantage	Disadvantage
(Cheng et al. 2021) [104] SwinTrack (Lin et al. 2021) [48] TransCenter (Xu et al. 2021) [118] (Zheng et al. 2021) [119] STASiam (Zhu et al. 2022) [120] Memot (Cai et al. 2022) [121] (Zhou et al. 2022) [53] (Chen et al. 2022) [122] AiATrack (Gao et al. 2022) [123]	superior feature extraction capability based on full attention and improve the ability of feature fusion utilizing dense multi-scale query to feature extraction and fusion utilizing the advantages of SOT to realize MOT multi-layer feature fusion and high versatility the fusion of long and short term features greatly improve the capability of data fusion incorporate attention module to enhance data fusion the “attention in attention” mechanism better captures the feature representation	high model complexity require a number of training models large memory consumption demand large resources complex model structure the increase in memory cost the size of the time window affects performance relatively slower performance on edge devices computationally costly
TCN (Pimpa et al. 2023) [101] LCA (He et al. 2023) [16] (Zhao et al. 2023) [124] TransTRDT (Sun et al. 2024) [17] CTMOT (Zhang et al. 2024) [38]	improve performance and reduce the complexity comprehensive data fusion of long-term frames and their contexts a flexible model possessing excellent feature-mixing capability exhibit adaptability to the complex environment integrate local and global features to improve performance	easily affected by other factors computationally costly — complex model structure the ambiguity in the accuracy of the data association stage
DAPNet (Zhu et al. 2019) [125] MDNet (Zhang et al. 2020) [103] LT-MDNet (Shao et al. 2021) [43] TBCCF (Wang et al. 2020) [126] ThrAtt (Wang et al. 2022) [105] MixFormer (Song et al. 2023) [51] EfficientFormer (Li et al. 2023) [75] TADN (Psalta et al. 2024) [127] PTQ4SAM (Lv et al. 2024) [128]	remove redundant and noisy features better tracking speed and performance, effectively handle fast moving targets robust and efficient for long-term tracking, effective in countering occlusion enhance the anti-interference capability and real-time performance enhance tracking performance and real-time capability significantly improves the real-time inference speed low latency, high performance, and fast inference speed on edge devices the end-to-end training improves efficiency and performance post-training quantization improves efficiency and ensures performance	complex model structure cannot strike a balance between accuracy and speed the performance in other environments is unknown the situation in other environments remains unknown incapable of handling fast motion and light changes high training cost unknown performance on other devices — the ambiguous cause of bimodal distribution

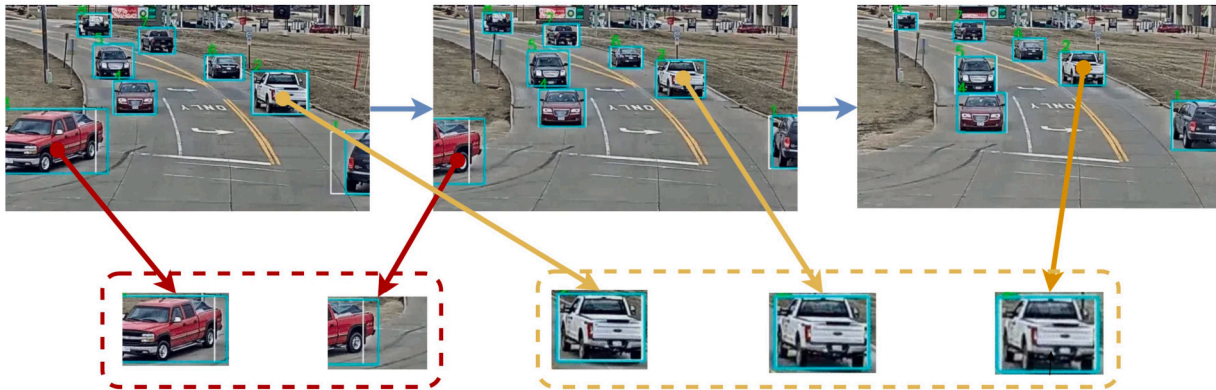


Fig. 7. Different types of object tracking data. The red and yellow dashed rectangles represent the spatial data of different tracking objects. The blue solid arrow represents temporal data. Space-time data requires a comprehensive consideration of both temporal and spatial data.

6.1.3. Spatio-temporal data fusion

The Global TRacking transformer (GTR) [53] is a transformer-based architecture designed to effectively address the challenge of object tracking data fusion. The GTR network takes a short sequence of frames as input and generates global trajectories for all detected objects. It encodes object features across all frames and utilizes trajectory queries to organize these features into coherent trajectories [53]. During the inference process, GTR processes the video stream using a sliding window approach. For each frame t , the image is fed into GTR to obtain the bounding box and object features for that frame, while maintaining a history buffer of t frames. In the initial frame, GTR initializes all detections as potential trajectories. For subsequent frame, GTR links the currently predicted trajectory to an existing one using the average assignment likelihood as a distance metric. The proposed method excels at efficiently integrating spatio-temporal data in object tracking.

6.2. Solutions for real-time inference

6.2.1. Quantization

PTQ4SAM [128] is a post-training quantization framework for the Segment Anything Model (SAM). It proposes a bimodal integration strategy that leverages equivalent mathematical symbolic operations to con-

vert the bimodal distribution into an offline normal distribution, which is relatively straightforward to quantify. Additionally, PTQ4SAM designs an adaptive granularity quantization method that focuses on quantizing the softmax operation in the attention mechanism. This model has achieved favorable outcomes in various vision tasks.

6.2.2. Prune

DAPNet [125] designs a dense aggregation module to acquire dense feature representation. Furthermore, it utilizes a feature pruning module to eliminate unimportant or rarely used weights in the model, thereby minimizing the influence of redundant features and interference information. This approach not only lightens the deep learning framework but also consistently enhances the inference speed for object tracking tasks.

6.2.3. Distillation

MixFormerV2 [51] introduces special prediction tokens that are mixed with the tokens of the target template and the search region. This enables it to rapidly capture the complex correlations between the target template and the search region. Additionally, MixFormerV2 adopts a simplified model based on distillation techniques [51], encompassing dense-to-sparse distillation and deep-to-shallow distillation to further

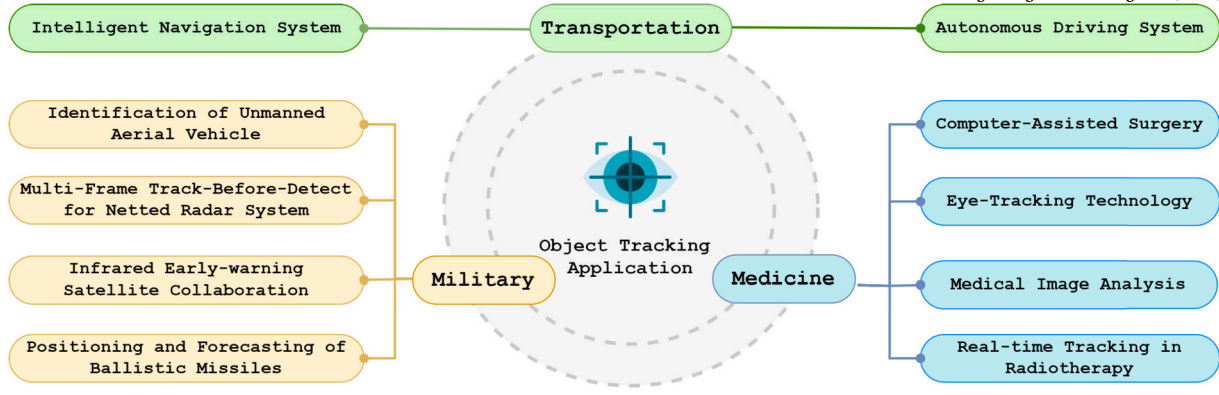


Fig. 8. The application of object tracking in military, transportation and medicine.

enhanced its real-time inference capabilities. Furthermore, this solution simplifies the object tracking process by eliminating custom convolutional classification and regression headers.

7. Solutions for environmental factors

Environmental factors pose challenges to object tracking, such as occlusion, illumination change, object deformation, object high-speed movement, which are persistent and prevalent obstacles in the field of object tracking. The majority of related research in object tracking addresses these environmental challenges.

7.1. Solutions based on deep learning

Addressing the challenges of object tracking in complex scenes, such as rotation, interference from similar objects, and occlusion, recent research [87] has introduced an object tracking algorithm based on Siamese CNN multi-scale rotation search, and an algorithm based on discrete correlation filter discriminant network [87]. These algorithms demonstrate impressive precision in boundary location and swift execution speeds, enabling them to tackle issues like occlusion and interference from similar objects in complex environments. Furthermore, other studies have been conducted based on adaptive convolutional features [102] and convolutional regression networks incorporating a fusion re-detection mechanism [95], both of which effectively address these tracking challenges.

Through enhancing the image quality during the object tracking process, the challenge posed by object occlusion and background clutter can be effectively addressed. For instance, MICU [44] integrates multi-level information compensation with the U-net network to enhance the quality of image super-resolution reconstruction, thereby alleviating the issue of background clutter in object tracking. Additionally, DNNAM [45] and MFMAM [129] leverage deep neural networks and the attention mechanism to more accurately infer the content of occluded regions in the image, which assists in mitigating object occlusion in object tracking tasks.

Regarding object tracking in low-light environments, denoising and low-light enhancement techniques can significantly enhance tracking performance [130]. For example, a recent study [76] proposes a object tracking system that incorporates hybrid denoising and low-light enhancement. This system is trained on a synthetic low-light dataset and demonstrates superior performance compared to the traditional Mix-Former [50]. This emphasizes the potential of appropriate preprocessing and enhancement methods to significantly improve the accuracy of object tracking in low-light conditions.

7.2. Solutions based on soft feature

Aiming at the problem of objects easily being lost due to occlusion, shape changes, and scale variations during object tracking, researchers

have devised a novel forward prediction tracking method based on Soft Feature (SF) [131]. This method involves selecting the targeted object area within the video image, filtering out scattered pixels, and marking the resulting pixel group. Subsequently, a time-domain trajectory is fitted, and soft features along with their constraint models are derived through computation. In addition, a recent study [132] utilizes an attention mechanism to conduct soft feature fusion between multiple pixels and points. This study introduces PointNet++ [133] to obtain multi-scale deep representations of point clouds, enabling adaptive interactive feature fusion between multi-scale soft features of images and point clouds. The SF exhibits robust anti-interference capabilities against changes in object shape and scale.

8. Application in military, transportation and medicine

Object tracking has an extremely wide range of applications and plays a pivotal role in many fields. In recent years, a significant number researchers have applied object tracking technology to various fields, resulting in numerous research outcomes, as shown in Fig. 8.

8.1. Military field

8.1.1. Positioning and forecasting of ballistic missiles

The space early warning system utilizes the angle of arrival measurement information of spaceborne infrared sensors to position and forecast the trajectories of ballistic missiles. By establishing both the active-phase motion model and the constrained nonlinear motion model for ballistic missiles, the system can effectively track and predict their movements [134].

8.1.2. Identification of unmanned aerial vehicle (UAV)

By developing a UAV object tracking and recognition system based on an intelligent pan-tilt-head system, accurate object tracking and recognition can be achieved through the employment of a consensus algorithm, fast image processing and multi-modal information fusion [135][136].

8.1.3. Infrared early-warning satellite collaboration

To enhance the accuracy of object tracking, some studies [137] have proposed cooperative object tracking methods under conditions of quantization noise. The motion model is established based on the motion features, aiming to improve the accuracy of position estimation.

8.1.4. Multi-frame track-before-detect (MF-TBD) for netted radar system (NRS)

The NRS confronts an increasingly complex electromagnetic environment, facing not only ground and sea clutter but also various electromagnetic interference threats [138]. Through MF-TBD, consecutive

frames of raw echo data are directly processed jointly, leveraging the object's motion features for multi-frame accumulation in the time dimension to improve the performance of the algorithm, thereby enhancing the monitoring effect of the NRS [138].

8.2. Transportation field

8.2.1. Intelligent navigation system

The navigation system can transform actions back to internal latent states by imitating the human navigation approach, forming a closed-loop system of perception, decision making and reasoning, thereby improving the adaptability and robustness of the navigation system to complex and dynamic real-world scenarios [139]. In complex environments, such as urban streets, the use of intelligent navigation encounters numerous challenges, including the identification and avoidance of dynamic obstacles, and the adaptation to environmental changes. To solve these problems, researchers have proposed a variety of solutions, for example, incorporating semantic information for object classification and tracking [140], and adopting multimodal data processing and trajectory planning [141].

8.2.2. Autonomous driving system

The development of driverless technology is aimed at improving driving safety, comfort and economy while lowering the traffic accident rate [6]. In recent years, autonomous driving has witnessed significant development. For instance, LiDAR-based 3D MOT is essential for comprehending continuous spatio-temporal 3D motion in the surrounding environment, which is of great significance for the perception layer [34]. The preview distance adaptive model based on the particle swarm optimization algorithm can improve the path tracking accuracy while considering turning frequency and vehicle stability [142]. Additionally, face recognition and autonomous driving are two closely related and continually evolving technological fields [143][144]. Specifically, algorithms such as OD-LBP [145], EDRM-LBP [146], NCDB-LBP [147], and TAO-LBP [148] have continually optimized and refined the Local Binary Pattern algorithm, significantly improving the accuracy of face recognition. This technological advancement has propelled not only the field of face recognition but also injected fresh impetus into the development of autonomous driving technology.

8.3. Medical field

8.3.1. Computer-assisted surgery

Electromagnetic tracking technology is one of the key technologies for realizing computer-assisted surgery. It can track small objects in real time and overcome the problem of line-of-sight limitation. Despite being extensively investigated, the development of this technique in clinical applications has been relatively sluggish [149]. The development of active infrared-based surgical mark tracking and positioning systems provides a new solution for surgical navigation [150].

8.3.2. Medical image analysis

The application of object tracking technology in medical image analysis is mainly manifested in the monitoring and analysis of dynamic processes. For instance, dynamic object tracking methods based on medical images can improve the tracking performance of moving objects, such as the material flow in the esophagus, thereby enabling the optimization of the entire swallowing process model [151]. The accuracy of superficial object contours can be significantly improved using endoscopic tracking and registration techniques, which is of great significance for the precise delineation of tumor volumes [152]. The MOT medical image detection method can effectively reduce the patient detection time and strive for the optimal treatment time for patients [153]. The multimodal imaging techniques in medical imaging analysis have provided an abundance of information for object definition in radiotherapy for head and neck cancer [154].

8.3.3. Eye-tracking technology

Eye-tracking technology is increasingly employed in the medical field, which can not only improve the accuracy of diagnosis but also improve medical performance and rehabilitation outcomes. This technique is utilized in numerous branches of medicine, including neurology, cardiology, pathology, surgery, rehabilitation, and other relevant fields, providing objective data of various medical conditions [30].

8.3.4. Real-time tracking in radiotherapy

In radiotherapy, for the treatment of moving targets, the uncertainty of the object position tends to cause treatment errors. Techniques for real-time imaging verification and tracking of moving objects, such as those based on implantable transponders and optical imaging, have been clinically implemented, significantly increasing the success rate of radiotherapy [155].

9. Discussion

Based on the comprehensive analysis of the challenges and corresponding solutions associated with object tracking, as previously outlined. The future trajectory and development directions of object tracking can be explored and discussed from the following four perspectives:

(1) Further Integration deep learning technology and improving interpretability. With the rapid advancement of deep learning technology, future object tracking will place greater emphasis on the integration of deep learning frameworks, particularly lightweight ones. This integration aims to enhance the performance and efficiency of tracking systems while maintaining interpretability.

(2) Object tracking combined with multi-modal fusion. In practical applications, object tracking often needs to handle multi-modal data, such as video, audio, and text. By fusing data from different modalities, richer and more comprehensive object information can be obtained, improving the accuracy and robustness of tracking systems.

(3) Edge devices typically possess constrained computing capabilities and storage resources. Meanwhile, real-time performance is a crucial criterion for the practical implementation of tracking systems. Therefore, enhancing the tracking speed on edge devices is a complex yet significant challenge that requires innovative solutions.

(4) Focus on practical applications and MOT. Future advancements in object tracking will increasingly emphasize research directed towards practical applications, especially MOT. To meet the actual needs of hot fields like autonomous driving and intelligent monitoring, more advanced object tracking algorithms and systems will be developed, ensuring higher accuracy, efficiency, and reliability.

10. Conclusion

In this paper, we provide a comprehensive overview of the related research on object tracking, encompassing the research background, related work, challenges and their underlying reasons, solutions, and applications across various domains. In recent years, numerous studies on object tracking have indicated that it remains a significant research direction in computer vision, with extensive utilization in numerous fields. Despite significant advancements in object tracking, the inherent challenges in this domain persist, necessitating continued exploration by researchers to develop superior solutions. We hope that this survey will serve as a practical reference for understanding the current research progress in object tracking, assisting readers in further enhancing object tracking for diverse innovations and applications.

Initially, this article intended to thoroughly examine the solutions pertaining to SOT and MOT. However, due to the complexity of the related content, it deviated slightly from my primary intention of focusing on challenges.

CRediT authorship contribution statement

Wenqi Zhang: Conceptualization, Methodology, Software, Writing – Original Draft, Investigation; **Xinqiang Li:** Writing – Reviewing and Editing; **Xingyu Liu:** Writing – Reviewing and Editing; **Shiteng Lu:** Investigation, Validation; **Huanling Tang:** Software, Conceptualization, Writing – Reviewing and Editing, Supervision, Project administration.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT, ERNIE Bot and other AI translation software in order to assist in the writing. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Huanling Tang reports financial support was provided by National Natural Science Foundation of China. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No.62341605, No.62372077). We are extremely grateful to Xiaoyan Liu and Zhenqi Wang for helpful discussions and feedback on initial versions of this paper. We would like to express a special thanks to ChatGPT, ERNIE Bot and other AI translation software, the advanced AI language model, for its assistance in providing clarity, coherence, and depth to my writing.

Data availability

No data was used for the research described in the article.

References

- [1] M. Shah, A. Yilmaz, O. Javed, Object tracking: a survey, *ACM Comput. Surv.* 38 (4) (2006), 13-es.
- [2] Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-net: convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18, Springer, 2015, pp. 234–241.
- [3] Xingxing Xie, Gong Cheng, Jiabao Wang, Xiwen Yao, Junwei Han, Oriented r-cnn for object detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 3520–3529.
- [4] Syed Sahil Abbas Zaidi, Mohammad Samar Ansari, Asra Aslam, Nadia Kanwal, Mamoon Asghar, Brian Lee, A survey of modern deep learning based object detection models, *Digit. Signal Process.* 126 (2022) 103514.
- [5] Justin Wilson, Ming C. Lin, Avot: audio-visual object tracking of multiple objects for robotics, in: 2020 IEEE International Conference on Robotics and Automation (ICRA), 2020, pp. 10045–10051.
- [6] Mingfei Cheng, Yuan Zhou, Xiaofei Xie, Behavexplor: behavior diversity guided testing for autonomous driving systems, in: Proceedings of the 32nd ACM SIGSOFT International Symposium on Software Testing and Analysis, 2023, pp. 488–500.
- [7] V.M. Opanasenko, Sh Kh Fazilov, S.S. Radjabov, Sh S. Kakharov, Multilevel face recognition system, *Cybern. Syst. Anal.* 60 (1) (2024) 146–151.
- [8] Xinyuan Qian, Qi Liu, Jiadong Wang, Haizhou Li, Three-dimensional speaker localization: audio-refined visual scaling factor estimation, *IEEE Signal Process. Lett.* 28 (2021) 1405–1409.
- [9] Matej Kristan, Jiri Matas, Ales Leonardis, Michael Felsberg, Luka Cehovin, Gustavo Fernandez, Tomas Vojir, Gustav Hager, Georg Nebel, Roman Pflugfelder, The visual object tracking vot2015 challenge results, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2015, pp. 1–23.
- [10] Pengxin Zeng, Peng Chen, Linlin Zhu, Tracking algorithm based on target motion model, *J. Syst. Simul.* 12 (2006) 3491–3494.
- [11] Mustansar Fiaz, Arif Mahmood, Sajid Javed, Soon Ki Jung, Handcrafted and deep trackers: recent visual object tracking approaches and trends, *ACM Comput. Surv.* 52 (2) (2019) 1–44.
- [12] Yuantao Chen, Runlong Xia, Kai Yang, Ke Zou, Image inpainting algorithm based on inference attention module and two-stage network, *Eng. Appl. Artif. Intell.* 137 (2024) 109181.
- [13] Zhen Cui, Shengtao Xiao, Jiashi Feng, Shuicheng Yan, Recurrently target-attending tracking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1449–1458.
- [14] Shuxiao Ding, Lukas Schneider, Marius Cordts, Juergen Gall, Ada-track: end-to-end multi-camera 3d multi-object tracking with alternating detection and association, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 15184–15194.
- [15] Ruopeng Gao, Limin Wang, Memotr: long-term memory-augmented transformer for multi-object tracking, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 9901–9910.
- [16] Kaijie He, Canlong Zhang, Sheng Xie, Zhixin Li, Zhiwen Wang, Target-aware tracking with long-term context attention, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, 2023, pp. 773–780.
- [17] Ziwen Sun, Lizhi Qian, Guanglin Yuan, Chuandong Yang, Chong Ling, Transformer object tracking method based on real-time dynamic template update, *Comput. Eng.* (2024) 1–10.
- [18] H. Terra Marco, Y. Ishihara João, S. Inoue Roberto, A new approach to robust linear filtering problems, *IFAC Proc. Vol.* 44 (1) (2011) 1174–1179.
- [19] Yuanyuan Fang, The target tracking based on cubature Kalman filter, *J. Acoust. Soc. Am.* 131 (4,Supplement) (2012) 3444.
- [20] Steven S. Beauchemin, John L. Barron, The computation of optical flow, *ACM Comput. Surv.* 27 (3) (1995) 433–466.
- [21] Ruibing Jin, Guosheng Lin, Changyun Wen, Jianliang Wang, Fayao Liu, Feature flow: in-network feature flow estimation for video object detection, *Pattern Recognit.* 122 (2022) 108323.
- [22] Efim Mazor, Amir Averbuch, Yakov Bar-Shalom, Joshua Dayan, Interacting multiple model methods in target tracking: a survey, *IEEE Trans. Aerosp. Electron. Syst.* 34 (1) (1998) 103–123.
- [23] Chee-Yee Chong, David Garren, Timothy P. Grayson, Ground target tracking—a historical perspective, in: 2000 IEEE Aerospace Conference. Proceedings (Cat. No. 00TH8484), vol. 3, IEEE, 2000, pp. 433–448.
- [24] Yifan Sun, Nicolas Bohm Agostini, Shi Dong, David Kaeli, Summarizing cpu and gpu design trends with product data, *arXiv preprint*, arXiv:1911.11313, 2019.
- [25] David Blythe, Rise of the graphics processor, *Proc. IEEE* 96 (5) (2008) 761–778.
- [26] R. Sukanya, K. Swaathikka, R. Soorya, Enhancing computational performance using cpu-gpu integration, *Int. J. Comput. Appl.* 111 (7) (2015).
- [27] David Luebke, Graphics hardware & gpu computing: past, present, and future, in: Proceedings of Graphics Interface 2009, 2009, p. 1.
- [28] Zihan Xie, Hongbin Gu, Dongsu Wu, Robust eye tracking technology based on multi-camera, *Meas. Control Technol.* 41 (12) (2022) 58–65.
- [29] Zhendong Song, Wei Jiang, Monocular multi-viewpoints stereo imaging and application, *Acta Opt. Sin.* 32 (05) (2012) 147–152.
- [30] Mohammed Tahri Sqalli, Begali Aslonov, Mukhammadjon Gafurov, Nurmukhammad Mukhammadiev, Yahya Sqalli Houssaini, Eye tracking technology in medical practice: a perspective on its diverse applications, *Front. Med. Technol.* 5 (2023) 1253001.
- [31] Ye Liu, 3d object tracking based on extended kernelized correlation filter, *J. Nanjing Univ. Posts Telecommun. Nat. Sci.* 38 (05) (2018) 79–84.
- [32] Weiyang Yin, Research on human motion-object detection and tracking algorithm-based on depth image, Master's thesis, Nanjing University of Aeronautics and Astronautics, 2020.
- [33] Vasyil Molebny, Paul McManamon, Ove Steinvall, Takao Kobayashi, Weibiao Chen, Laser radar: historical prospective—from the East to the West, *Opt. Eng.* 56 (3) (2017) 031220.
- [34] Zhenkai Xiong, Xiaoqiang Cheng, Youdong Wu, Zhiqiang Zuo, Jiasheng Liu, Lidar-based 3d multi-object tracking for unmanned vehicles, *Acta Autom. Sin.* 49 (10) (2023) 2073–2083.
- [35] Xi Li, Yufei Zha, Tianzhu Zhang, Zhen Cui, Wangmeng Zuo, Zhiqiang Hou, Huchuan Lu, Hanzhi Wang, Survey of visual object tracking algorithms based on deep learning, *J. Image Graph.* 24 (12) (2019) 2057–2080.
- [36] A. Vaswani, Attention is all you need, *Adv. Neural Inf. Process. Syst.* (2017).
- [37] Shoufa Chen, Peize Sun, Yibing Song, Ping Luo, Diffusionet: diffusion model for object detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 19830–19843.
- [38] Yingjun Zhang, Xiaohui Bai, Binhong Xie, Multi-object tracking algorithm based on cnn-transformer feature fusion, *Comput. Eng. Appl.* 60 (02) (2024) 180–190.
- [39] Puja Bharati, Ankita Pramanik, Deep learning techniques—r-cnn to mask r-cnn: a survey, in: Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019, 2020, pp. 657–668.
- [40] J. Redmon, You only look once: unified, real-time object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.

- [41] Muhammad Hussain, Yolo-v1 to yolo-v8, the rise of yolo and its complementary nature toward digital manufacturing and industrial defect detection, *Machines* 11 (7) (2023) 677.
- [42] Tianheng Cheng, Lin Song, Yixiao Ge, Wenyu Liu, Xinggang Wang, Ying Shan, Yolo-world: real-time open-vocabulary object detection, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 16901–16911.
- [43] Jiangnan Shao, Hongwei Ge, A long-term object tracking algorithm based on deep learning and object detection, *CAAI Trans. Intell. Syst.* 16 (03) (2021) 433–441.
- [44] Yuantao Chen, Runlong Xia, Yang Ke Zou, Micu: image super-resolution via multi-level information compensation and u-net, *Expert Syst. Appl.* 245 (Jul.) (2024) 123111.1–123111.9.
- [45] Yuantao Chen, Runlong Xia, Kai Yang, Ke Zou, Dnnam: image inpainting algorithm via deep neural networks and attention mechanism, *Appl. Soft Comput.* 154 (2024) 111392.
- [46] Xin Chen, Bin Yan, Jiawen Zhu, Dong Wang, Xiaoyun Yang, Huchuan Lu, Transformer tracking, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8126–8135.
- [47] Bin Yan, Houwen Peng, Jianlong Fu, Dong Wang, Huchuan Lu, Learning spatiotemporal transformer for visual tracking, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10448–10457.
- [48] Liting Lin, Heng Fan, Zhipeng Zhang, Yong Xu, Haibin Ling, Swintrack: a simple and strong baseline for transformer tracking, *Adv. Neural Inf. Process. Syst.* 35 (2022) 16743–16754.
- [49] Christoph Mayer, Martin Danelljan, Goutam Bhat, Matthieu Paul, Danda Pani Paudel, Fisher Yu, Luc Van Gool, Transforming model prediction for tracking, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 8731–8740.
- [50] Yutao Cui, Cheng Jiang, Limin Wang, Gangshan Wu, Mixformer: end-to-end tracking with iterative mixed attention, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 13608–13618.
- [51] Yutao Cui, Tianhui Song, Gangshan Wu, Limin Wang, Mixformerv2: efficient fully transformer tracking, *Adv. Neural Inf. Process. Syst.* 36 (2024).
- [52] Peize Sun, Jinkun Cao, Yi Jiang, Rufeng Zhang, Enze Xie, Zehuan Yuan, Changhu Wang, Ping Luo, Transtrack: multiple object tracking with transformer, *arXiv preprint, arXiv:2012.15460*, 2020.
- [53] Xingyi Zhou, Tianwei Yin, Vladlen Koltun, Philipp Krähenbühl, Global tracking transformers, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 8771–8780.
- [54] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, Xinggang Wang, Bytetrack: multi-object tracking by associating every detection box, in: *European Conference on Computer Vision*, Springer, 2022, pp. 1–21.
- [55] Fan Yang, Shigeyuki Odashima, Shoichi Masui, Shan Jiang, Hard to track objects with irregular motions and similar appearances? Make it easier by buffering the matching space, in: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 4799–4808.
- [56] Jinkun Cao, Jiangmiao Pang, Xishuo Weng, Rawal Khiradkar, Kris Kitani, Observation-centric sort: rethinking sort for robust multi-object tracking, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 9686–9696.
- [57] Yu-Hsiang Wang, Jun-Wei Hsieh, Ping-Yang Chen, Ming-Ching Chang, Hung-Hin So, Xin Li, Smiletrack: similarity learning for occlusion-aware multiple object tracking, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, 2024, pp. 5740–5748.
- [58] Changcheng Xiao, Qiong Cao, Yujie Zhong, Long Lan, Xiang Zhang, Zhigang Luo, Dacheng Tao, Motiontrack: learning motion predictor for multiple object tracking, *Neural Netw.* 179 (2024) 106539.
- [59] Tao Gao, Zhengguang Liu, Jun Zhang, Shihong Yue, Feature points based multiple moving targets tracking, *J. Electron. Inf. Technol.* 32 (05) (2010) 1111–1115.
- [60] Huayu Su, Key Techniques Research on GPU Parallel Computing Targeted on Applications, PhD thesis, National University of Defense Technology, 2014.
- [61] Hongtao Bai, Research on High performance Parallel Algorithms based on GPU, PhD thesis, Jilin University, 2010.
- [62] Qiao Yan, Research on moving object tracking based on multi-camera, Master's thesis, Beijing Jiaotong University, 2020.
- [63] Kolja Kuhnlenz, Mathias Bachmayer, Martin Buss, A multi-focal high-performance vision system, in: *Proceedings 2006 IEEE International Conference on Robotics and Automation*, ICRA 2006, IEEE, 2006, pp. 150–155.
- [64] Huishu Ke, Research on 3d reconstruction-technology with depth cameras, Master's thesis, Southeast University, 2019.
- [65] Zijian Zhao, Research on image processing based on depth camera, Master's thesis, University of Science and Technology of China, 2022.
- [66] Chenxu Luo, Xiaodong Yang, Alan Yuille, Exploring simple 3d multi-object tracking for autonomous driving, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10488–10497.
- [67] Guoyan Xu, Huan Niu, Chenyang Guo, Hongjie Su, Research on target recognition and tracking based on 3d laser point cloud, *Automot. Eng.* 42 (01) (2020) 38–46.
- [68] Gioele Ciaparrone, Francisco Duque Sánchez, Siham Tabik, Luigi Troiano, Roberto Tagliaferri, Francisco Herrera, Deep learning in video multi-object tracking: a survey, *Neurocomputing* 381 (2020) 61–88.
- [69] Xiankai Lu, Research on Object Tracking Based on Deep Learning, PhD thesis, Shanghai Jiaotong University, 2018.
- [70] Mengjie Hu, Xiaotong Zhu, Haotian Wang, Shixiang Cao, Chun Liu, Qing Song, Stdformer: spatial-temporal motion transformer for multiple object tracking, *IEEE Trans. Circuits Syst. Video Technol.* 33 (11) (2023) 6571–6594.
- [71] Hongbin Liu, Detection Based Data Association Method for Multi-Target Tracking in Complex Scene, PhD thesis, Shandong University, 2019.
- [72] Run Luo, Zikai Song, Lintao Ma, Jinlin Wei, Wei Yang, Min Yang, Diffusiontrack: diffusion model for multi-object tracking, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, 2024, pp. 3991–3999.
- [73] Longqiang Ni, Research on Key Technologies for Data Processing of Target Tracking System, PhD thesis, Northwestern Polytechnical University, 2016.
- [74] Kai Huang, Jun Chu, Lu Leng, Xingbo Dong, Tatrack: target-aware transformer for object tracking, *Eng. Appl. Artif. Intell.* 127 (2024) 107304.
- [75] Yanyu Li, Ju Hu, Yang Wen, Georgios Evangelidis, Kamyar Salahi, Yanzhi Wang, Sergey Tuluyakov, Jian Ren, Rethinking vision transformers for mobilenet size and speed, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 16889–16900.
- [76] Anqi Yi, Nantheera Anantrasirchai, A comprehensive study of object tracking in low-light environments, *Sensors* 24 (13) (2024) 4359.
- [77] Chi Feng, Xiaofeng Lv, Qingbo Ji, Particle filtering theory and its application in target tracking, *Comput. Eng. Appl.* 06 (2008) 246–248.
- [78] M. Sanjeev Arulampalam, Simon Maskell, Neil Gordon, Tim Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, *IEEE Trans. Signal Process.* 50 (2) (2002) 174–188.
- [79] Yuxia Wang, Qingjie Zhao, Yiming Cai, Bo Wang, Tracking by auto-reconstructing particle filter trackers, *Chinese J. Comput.* 39 (07) (2016) 1294–1306.
- [80] Xu Wang, Yi Liu, Guoyan Li, Target tracking algorithm based on improved kernelized correlation filters, *J. Chin. Comput. Syst.* 41 (03) (2020) 506–510.
- [81] Liyang Yu, Chunxiao Fan, Yue Ming, Improved target tracking algorithm based on kernelized correlation filter, *J. Comput. Appl.* 35 (12) (2015) 3550–3554.
- [82] Sergio E. Ontiveros-Gallardo, Vitaly Kober, Objects tracking with adaptive correlation filters and Kalman filtering, in: Abdul A.S. Awwal, Khan M. Iftikharuddin, Mohammad A. Matin, Mireya García Vázquez, Andrés Márquez (Eds.), *Optics and Photonics for Information Processing IX*, vol. 9598, International Society for Optics and Photonics, SPIE, 2015, p. 95980X.
- [83] Cory Miller, Bethany Allik, Mark Ilg, Ryan Zurakowski, Kalman filter-based tracking of multiple similar objects from a moving camera platform, in: *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, 2012, pp. 5679–5684.
- [84] Takahiro Kodama, T. Yamaguchi, Hiroshi Harada, A method of object tracking based on particle filter and optical flow, in: *2009 ICCAS-SICE*, 2009, pp. 2685–2690.
- [85] Florian Deprez, Vladan Popovic, Beat Ott, Peter Wellig, Yusuf Leblebici, Real-time object detection and tracking in omni-directional surveillance using gpu, in: *Target and Background Signatures*, vol. 9653, SPIE, 2015, pp. 224–236.
- [86] Yan Yu, Target recognition and location based on monocular vision, Master's thesis, Harbin Engineering University, 2017.
- [87] Yuan Liu, Research on object tracking under complex-scene based on convolutional neural networks, PhD thesis, Nanjing University of Science and Technology, 2021.
- [88] Jiongning Sun, Taizhi Lv, Juan Zhang, Haitao Guo, Target tracking method based on deep learning and spatiotemporal prediction, *Radioengineering* 51 (09) (2021) 909–914.
- [89] Gabin Schieffer, Nattawat Pornthisan, Daniel Medeiros, Stefano Markidis, Jacob Wahlgren, Ivy Peng, Boosting the performance of object tracking with a half-precision particle filter on gpu, in: *European Conference on Parallel Processing*, Springer, 2023, pp. 294–305.
- [90] Qing Guo, Wei Feng, Ce Zhou, Rui Huang, Liang Wan, Song Wang, Learning dynamic Siamese network for visual object tracking, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1763–1771.
- [91] Junwei Li, Xiaolong Zhou, Sixian Chan, Shengyong Chen, A novel video target tracking method based on adaptive convolutional neural-network feature, *J. Comput.-Aided Des. Comput. Graph.* 30 (02) (2018) 273–281.
- [92] Jin Wu, Guohao Dong, Qiaoshen Li, Object tracking based on region convolution neural network and optical flow method, *Telecommun. Eng.* 58 (01) (2018) 6–12.
- [93] Jun Miao, Kai Li, Shaowu Xu, Target tracking based on multi-layer feature fusion of convolutional neural network, *Mod. Electron. Tech.* 41 (24) (2018) 114–116.
- [94] Xin Li, Chao Ma, Baoyuan Wu, Zhenyu He, Ming-Hsuan Yang, Target-aware deep tracking, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1369–1378.
- [95] Yongchao Jia, Xiaowei He, Zhonglong Zheng, Object tracking algorithm combining re-detection mechanism and convolutional regression network, *J. Comput. Appl.* 39 (08) (2019) 2247–2251.
- [96] Jianpeng Li, Zhenhong Shang, Hui Liu, Visual object tracking algorithm based on correlation filters with hierarchical convolutional features, *Comput. Sci.* 46 (07) (2019) 252–257.
- [97] Haibo Zhang, Target tracking algorithm based on regularized convolution neural network, *Inf. Technol.* 43 (06) (2019) 82–86, +90.
- [98] Zhongdao Wang, Hengshuang Zhao, Ya-Li Li, Shengjin Wang, Philip Torr, Luca Bertinetto, Do different tracking tasks require different appearance models?, *Adv. Neural Inf. Process. Syst.* 34 (2021) 726–738.

- [99] Fan Ma, Mike Zheng Shou, Linchao Zhu, Haoqi Fan, Yilei Xu, Yi Yang, Zhicheng Yan, Unified transformer tracker for object tracking, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 8781–8790.
- [100] Qi Chu, Wanli Ouyang, Hongsheng Li, Xiaogang Wang, Bin Liu, Nenghai Yu, Online multi-object tracking using cnn-based single object tracker with spatial-temporal attention mechanism, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4836–4845.
- [101] Pimpa Cheewaparakobkit, Chih-Yang Lin, Timothy K. Shih, Avirmed Enkhbat, Enhancing single object tracking with a hybrid approach: temporal convolutional networks, attention mechanisms, and spatial-temporal memory, IEEE Access (2023).
- [102] Yuzhu Cai, Dedong Yang, Ning Mao, Fucui Yang, Visual tracking algorithm based on adaptive convolutional features, Acta Opt. Sin. 37 (03) (2017) 269–280.
- [103] Xiaoli Zhang, Longxin Zhang, Mansheng Xiao, Guocai Zuo, Target tracking by deep fusion of fast multi-domain convolutional neural network and optical flow method, Comput. Eng. Sci. 42 (12) (2020) 2217–2222.
- [104] Xu Cheng, Yiping Cui, Chen Song, Beijing Chen, Yuhui Zhen, Jingang Shi, Object tracking algorithm based on temporal-spatial attention mechanism, Comput. Sci. 48 (04) (2021) 123–129.
- [105] Ling Wang, Lei Zhou, Peng Wang, Yane Bai, Research on tracking algorithm based on attention mechanism and Siamese network, Comput. Eng. Appl. 58 (23) (2022) 161–168.
- [106] Yi Wu, Jongwoo Lim, Ming-Hsuan Yang, Online object tracking: a benchmark, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 2411–2418.
- [107] Giorgio Roffo, Simone Melzi, et al., The visual object tracking vot2016 challenge results, in: Computer Vision—ECCV 2016 Workshops: Amsterdam, the Netherlands, October 8–10 and 15–16, 2016, Proceedings, Part II, Springer International Publishing, 2016, pp. 777–823.
- [108] UT Benchmark, A benchmark and simulator for uav tracking, in: European Conference on Computer Vision, 2016.
- [109] Matthias Muller, Adel Bibi, Silvio Giancola, Salman Alsubaihi, Bernard Ghanem, Trackingnet: a large-scale dataset and benchmark for object tracking in the wild, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 300–317.
- [110] Lianghua Huang, Xin Zhao, Kaiqi Huang, Got-10k: a large high-diversity benchmark for generic object tracking in the wild, IEEE Trans. Pattern Anal. Mach. Intell. 43 (5) (2019) 1562–1577.
- [111] Heng Fan, Liting Lin, Fan Yang, Peng Chu, Ge Deng, Sijia Yu, Hexin Bai, Yong Xu, Chunyuan Liao, Haibin Ling, Lasot: a high-quality benchmark for large-scale single object tracking, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 5374–5383.
- [112] L. Leal-Taixé, Motchallenge 2015: towards a benchmark for multi-target tracking, arXiv preprint, arXiv:1504.01942, 2015.
- [113] Kevin Lai, Liefeng Bo, Xiaofeng Ren, Dieter Fox, A large-scale hierarchical multi-view rgb-d object dataset, in: 2011 IEEE International Conference on Robotics and Automation, IEEE, 2011, pp. 1817–1824.
- [114] Pushmeet Kohli Nathan Silberman, Derek Hoiem, Rob Fergus, Indoor segmentation and support inference from rgbd images, in: ECCV, 2012.
- [115] P. Dendorfer, Mot20: a benchmark for multi object tracking in crowded scenes, arXiv preprint, arXiv:2003.09003, 2020.
- [116] Achal Dave, Tarasha Khurana, Pavel Tokmakov, Cordelia Schmid, Deva Ramanan, Tao: a large-scale benchmark for tracking any object, in: Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16, Springer, 2020, pp. 436–454.
- [117] Peize Sun, Jinkun Cao, Yi Jiang, Zehuan Yuan, Song Bai, Kris Kitani, Ping Luo, Dancetrack: multi-object tracking in uniform appearance and diverse motion, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 20993–21002.
- [118] Xu Yihong, Ban Yutong, Delorme Guillaume, Gan Chuang, Rus Daniela, Alameda-Pineda Xavier, Transcenter: transformers with dense representations for multiple-object tracking, IEEE Trans. Pattern Anal. Mach. Intell. (2022), PP.
- [119] Linyu Zheng, Ming Tang, Yingying Chen, Guibo Zhu, Jinqiao Wang, Hanqing Lu, Improving multiple object tracking with single object tracking, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 2453–2462.
- [120] Wenqiu Zhu, Guang Zou, Zhigao Zeng, Xiaoyi Wang, Target tracking fusing spatio-temporal contextual information and attention mechanism, Comput. Eng. Des. 43 (09) (2022) 2567–2577.
- [121] Jiarui Cai, Mingze Xu, Wei Li, Yuanjun Xiong, Wei Xia, Zhuowen Tu, Stefano Soatto, Memot: multi-object tracking with memory, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 8090–8100.
- [122] Xin Chen, Bin Yan, Jiawen Zhu, Huchuan Lu, Xiang Ruan, Dong Wang, High-performance transformer tracking, IEEE Trans. Pattern Anal. Mach. Intell. 45 (7) (2022) 8507–8523.
- [123] Shenyan Gao, Chunlun Zhou, Chao Ma, Xinggang Wang, Junsong Yuan, Aiattrack: attention in attention for transformer visual tracking, in: European Conference on Computer Vision, Springer, 2022, pp. 146–164.
- [124] Jie Zhao, Johan Edstedt, Michael Felsberg, Dong Wang, Huchuan Lu, Leveraging the power of data augmentation for transformer-based tracking, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2024, pp. 6469–6478.
- [125] Yabin Zhu, Chenglong Li, Bin Luo, Jin Tang, Xiao Wang, Dense feature aggregation and pruning for rgbt tracking, in: Proceedings of the 27th ACM International Conference on Multimedia, 2019, pp. 465–472.
- [126] Sikui Wang, Yunpeng Liu, Lin Qi, Zhongyu Zhang, Zhiyuan Liu, Object tracking method based on background constraints and convolutional features, Comput. Eng. Appl. 56 (08) (2020) 205–214.
- [127] Athena Psalta, Vasileios Tsironis, Konstantinos Karantzas, Transformer-based assignment decision network for multiple object tracking, Comput. Vis. Image Underst. 241 (2024) 103957.
- [128] Chengtao Lv, Hong Chen, Jinyang Guo, Yifu Ding, Xianglong Liu, Ptt4sam: post-training quantization for segment anything, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 15941–15951.
- [129] Yuantao Chen, Runlong Xia, Kai Yang, Ke Zou, Mfmam: image inpainting via multi-scale feature module with attention module, Comput. Vis. Image Underst. 238 (2024) 103883.
- [130] Shekhar Karanwal, Manoj Diwakar, Enhanced lpq based two novel blur invariant face descriptors in light variations, in: International Conference on Soft Computing and Pattern Recognition, Springer, 2021, pp. 156–169.
- [131] Wentao Jiang, Wanjuan Liu, Heng Yuan, Research of object tracking based on soft feature theory, Chinese J. Comput. 39 (07) (2016) 1334–1355.
- [132] Guangming Wang, Chensheng Peng, Yingying Gu, Jinpeng Zhang, Hesheng Wang, Interactive multi-scale fusion of 2d and 3d features for multi-object vehicle tracking, IEEE Trans. Intell. Transp. Syst. 24 (10) (2023) 10618–10627.
- [133] Charles Ruizhongtai Qi, Li Yi, Hao Su, Leonidas J. Guibas, Pointnet++: deep hierarchical feature learning on point sets in a metric space, Adv. Neural Inf. Process. Syst. 30 (2017).
- [134] Dun Li, Target location and forecast in space early warning system, PhD thesis, National University of Defense Technology, 2001.
- [135] Shuaijun Wang, Fan Jiang, Bin Zhang, Rui Ma, Qi Hao, Development of uav-based target tracking and recognition systems, IEEE Trans. Intell. Transp. Syst. 21 (8) (2019) 3409–3422.
- [136] Ziang Cao, Ziyuan Huang, Liang Pan, Shiwei Zhang, Ziwei Liu, Changhong Fu, Tctrack: temporal contexts for aerial tracking, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 14798–14808.
- [137] Yu Wang, Cheng Ma, Bo Peng, Jiuliang Tao, Yifan Yu, Bo Zhao, High precision co-operation target tracking method for space-based infrared early-warning satellite, J. Chin. Inertial Technol. 30 (02) (2022) 257–263.
- [138] Feilong Ku, Research on multi-frame tracking before detection-technology on airborne early warning radars, Master's thesis, University of Electronic Science and Technology of China, 2020.
- [139] Wenqi Zhang, Kai Zhao, Peng Li, Xiao Zhu, Yongliang Shen, Yanna Ma, Yingfeng Chen, Weiming Lu, A closed-loop perception, decision-making and reasoning mechanism for human-like navigation, arXiv preprint, arXiv:2207.11901, 2022.
- [140] Andreas Ess, Konrad Schindler, Bastian Leibe, Luc Van Gool, Object detection and tracking for autonomous navigation in dynamic environments, Int. J. Robot. Res. 29 (14) (2010) 1707–1725.
- [141] Xiafei Huang, Jing Liang, Xiaofeng Shen, Qilian Liang, A multimodal data harness approach of mobile sensors trajectory planning for target tracking, IEEE Internet Things J. 10 (11) (2022) 9252–9261.
- [142] Zhiguo Zhao, Liangjie Zhou, Qiang Zhu, Preview distance adaptive optimization for the path tracking-control of unmanned vehicle, J. Mech. Eng. 54 (24) (2018) 166–173.
- [143] Shekhar Karanwal, Manoj Diwakar, Two novel color local descriptors for face recognition, Optik 226 (2021) 166007.
- [144] Anjith George, Christophe Ecabert, Hatem Otroushi Shahreza, Ketan Kotwal, Sébastien Marcel, Edgeface: efficient face recognition model for edge devices, IEEE Trans. Biom. Behav. Identity Sci. (2024).
- [145] Shekhar Karanwal, Manoj Diwakar, Od-lbp: orthogonal difference-local binary pattern for face recognition, Digit. Signal Process. 110 (2021) 102948.
- [146] Shekhar Karanwal, Manoj Diwakar, Edrm-lbp: effective directional radial median local binary pattern for face recognition, Int. J. Embed. Syst. 15 (6) (2022) 475–492.
- [147] Shekhar Karanwal, Manoj Diwakar, Neighborhood and center difference-based-lbp for face recognition, Pattern Anal. Appl. 24 (2021) 741–761.
- [148] Shekhar Karanwal, Manoj Diwakar, Triangle and orthogonal local binary pattern for face recognition, Multimed. Tools Appl. 82 (23) (2023) 36179–36205.
- [149] Alfred M. Franz, Tamas Haidegger, Wolfgang Birkfellner, Kevin Cleary, Terry M. Peters, Lena Maier-Hein, Electromagnetic tracking in medicine—a review of technology, validation, and applications, IEEE Trans. Med. Imaging 33 (8) (2014) 1702–1725.
- [150] Thore Kaser, Florian Niebling, Rahel Schmied-Kowarz, Rebecca Rodeck, Gerko Wende, Exploring requirements for neurosurgical augmented reality: design and evaluation of an infrared-based inside-out tracking approach on hololens 2, in: International Conference on Human-Computer Interaction, Springer, 2024, pp. 162–177.
- [151] Guofeng Qin, Jiahao Qin, Qiufang Xia, Jianghuang Zou, Pengpeng Lin, Chengkun Ren, Ruihan Wang, Dynamic target tracking method based on medical imaging, Front. Physiol. 13 (2022) 894282.
- [152] Brady Hunt, Gobind S. Gill, Daniel A. Alexander, Samuel S. Streeter, David J. Gladstone, Gregory A. Russo, Bassem I. Zaki, Brian W. Pogue, Rongxiao Zhang, Fast

deformable image registration for real-time target tracking during radiation therapy using cine mri and deep learning, *Int. J. Radiat. Oncol. Biol. Phys.* 115 (4) (2023) 983–993.

[153] Feng Long, Medical image analysis based on multi-object tracking, *World Latest Med. Inf.* 18 (19) (2018) 160–161.

[154] Kenneth Jensen, Gina Al-Farra, Danijela Dejanovic, Jesper G. Eriksen, Annika Loft, Christian R. Hansen, Frank A. Pameijer, Ruta Zukauskaitė, Cai Grau, Imaging for target delineation in head and neck cancer radiotherapy, in: *Seminars in Nuclear Medicine*, vol. 51, Elsevier, 2021, pp. 59–67.

[155] Elia Lombardo, Jennifer Dhont, Denis Page, Cristina Garibaldi, Real-time motion management in mri-guided radiotherapy: current status and ai-enabled prospects, *Radiother. Oncol.* 190 (2024) 109970.



Wenqi Zhang is a M.S. in the School of Computer Science and Technology at Shandong Technology and Business University. His research interests include Machine Learning, Object Tracking, Model Lightweight, and other related fields.



Xinqiang Li received his bachelor's degree from Shandong Technology and Business University. He is now a research and development engineer of Internet of Things Business Department of Smart Energy Branch of Shandong Luruan Digital Technology Co., LTD. His research interests include Machine Learning, Artificial Intelligence, and Data Mining.



Digital Signal Processing 161 (2025) 105082

Xingyu Liu is a M.S. in the School of Southeast University - Monash University Joint Graduate School (Suzhou). Her research interests include Artificial Intelligence Agents, Large Language Models, and other related domains.



Shiteng Lu is a M.S. in the School of Computer Science and Technology at Shandong Technology and Business University. His research interests include Multi-modal Knowledge Graph and Deep Learning.



Huanling Tang received the B.S. in Yantai University in 1993, the M.S. degree from Tsinghua University in 2004, and Ph.D. degree from Dalian Maritime University in 2009. Now she is a professor in the School of Computer Science and Technology at Shandong Technology and Business University. Her research interests include Machine Learning, Artificial Intelligence and Data Mining.