2025 International Conference on Economics, Management and Education Technology (ICEMET 2025)

Article

A Real-time Object Tracking Strategy for a Mobile Terminal Based on a Lightweight Visual Model

Bohan Qiu [1] and Ao Liu [2,*]

[1] School of Information Technology, Beijing Institute of Technology Zhuhai, Zhuhai, China
[2] College of Science, Mathematics and Technology, Wenzhou-Kean University, Wenzhou, China
* Correspondence: Ao Liu, College of Science, Mathematics and Technology, Wenzhou-Kean University, Wenzhou, China

Abstract: Due to the increasing popularity of mobile intelligent terminals and the rapid advancement of computer vision technology, real-time object tracking technology on mobile devices has become a demand in fields such as augmented reality, intelligent security, and unmanned systems. Target tracking algorithms often have difficulties in achieving efficient and real-time performance on mobile devices due to limitations such as limited computing resources and high power consumption. Therefore, it is urgent to improve its comprehensive performance through a lightweight visual model and an optimization strategy. This paper focuses on the application of a lightweight visual model in mobile real-time object tracking. Firstly, it analyzes the basic process of object tracking and its practical application scenarios, and introduces the structural characteristics and technical advantages of various lightweight models. To address the challenges of limited computing resources, the need for real-time performance, and the ability to adapt to varying scene complexities when deployed on mobile devices, a feature extraction approach is introduced that combines multi-scale, lightweight feature fusion with a dynamic adjustment system. It will improve the model's efficiency and performance when running on mobile devices. Additionally, to address the obstacles caused by illumination changes and target occlusion in complex scenes, an adaptive feature calibration method and a tracking state correction mechanism are proposed, which significantly enhance the system's robustness and tracking accuracy. The experimental results demonstrate that the proposed strategy enables efficient and stable real-time object tracking on mobile devices, providing a feasible technical path for applications.

Keywords: object tracking; lightweight models; mobile devices; feature fusion; computer vision

Received: 04 February 2026
Revised: 23 March 2026
Accepted: 04 April 2026
Published: 11 April 2026

1. Introduction

1.1. Research Background

The rapid development of Internet of Things (IoT) devices and mobile smart terminals, such as smartphones and drones, has made real-time object tracking technology an essential requirement in fields like intelligent surveillance, autonomous driving, and augmented reality. Traditional deep learning vision models, including those designed for high accuracy, demonstrate exceptional performance in terms of precision. However, their high computational complexity and large parameter volume pose significant challenges for mobile devices, which are constrained by limited computational resources and strict power consumption requirements. Mobile devices, often equipped with low-power processors such as those in the ARM Cortex-A series and limited memory bandwidth, frequently encounter increased latency or significant power consumption spikes when processing high-resolution video streams [1]. In practical applications, additional challenges such as variations in lighting conditions, occlusion, and rapid motion further compromise the robustness and reliability of tracking algorithms. In recent years, lightweight vision models, such as those employing depthwise separable convolutions, have effectively reduced computational burdens. These advancements have opened new possibilities for achieving real-time object tracking on mobile devices, offering a promising direction for addressing the limitations of traditional models.

Nevertheless, these approaches still face a fundamental trade-off between model compression and adaptability to diverse scenarios. Excessive compression can significantly degrade the model's ability to extract critical features, thereby reducing its overall effectiveness in complex environments. On the other hand, the introduction of advanced optimization techniques, such as dynamic template updates and multi-scale feature fusion, often results in increased computational latency, which can hinder real-time performance. This creates a challenging dilemma for researchers aiming to optimize tracking systems for resource-constrained environments [2]. Achieving an optimal balance between tracking accuracy, real-time responsiveness, and energy efficiency remains a critical focus of ongoing research. Addressing this challenge requires innovative strategies that can simultaneously enhance the model's adaptability to diverse conditions while maintaining computational efficiency and minimizing energy consumption. Such advancements are essential for enabling the widespread deployment of real-time object tracking technologies in practical applications, particularly on mobile platforms with limited resources.

1.2. Research Significance

Object tracking technology is advancing toward greater efficiency and intelligence. From a theoretical perspective, this research contributes to the enhancement of the existing theoretical framework in the field of computer vision. By focusing on the application of lightweight visual models in mobile real-time object tracking scenarios, the study delves into innovative methods such as multi-scale lightweight feature fusion, model pruning, and quantification. These approaches provide fresh theoretical insights and technical pathways for optimizing object tracking algorithms, thereby addressing current limitations and paving the way for future advancements in the field [3].

From a practical application perspective, the findings of this research hold substantial value. In the domain of intelligent security, a mobile real-time object tracking system based on lightweight visual models can enable immediate tracking and alerting of objects within monitored areas, significantly improving the efficiency and accuracy of security systems. In intelligent traffic scenarios, this technology facilitates real-time tracking of vehicles and pedestrians, offering critical support for traffic management and autonomous driving systems. Furthermore, in augmented reality applications, it enables real-time tracking of targets within real-world environments, enhancing user interaction and experience. Overall, this research effectively reduces the computational burden and energy consumption of mobile devices, extends battery life, and enhances device performance [2]. These improvements are expected to accelerate the widespread adoption of mobile real-time object tracking technology across various industries, fostering innovation and practical utility in diverse fields.

2. Overview of Related Technologies

2.1. Object Tracking Technology

2.1.1. The Basic Concept of Object Tracking

Object tracking is a technique that continuously identifies and monitors a specific object within a video sequence [4]. Its fundamental principle involves analyzing the characteristics of the object across consecutive images to consistently update its position and size. This technology enables the determination of the movement trajectory of a target by examining a sequence of images. The basic process includes four key steps: initialization, where the initial position and attributes of the target are identified through detection or manual marking in the first frame; feature extraction, which involves obtaining semantic information such as texture and shape using convolutional networks or lightweight models; state prediction, where the movement and size variations of the target are forecasted using methods like optical flow or correlation filtering algorithms; and model update, which dynamically adjusts the template to accommodate changes such as deformation or occlusion. These steps collectively ensure the accurate tracking of objects in dynamic environments.

The LightTrack model integrates neural architecture search to optimize the tracking process while compressing the single-stage network responsible for feature fusion and location regression. This approach significantly reduces computational complexity, making it more compatible with hardware constraints. Unlike traditional methods, this lightweight model prioritizes efficiency and adaptability. It employs a multi-scale feature pyramid to improve the detection of small targets, enhancing precision in scenarios with limited visibility or intricate details [5]. Additionally, quantitative perception training is utilized to minimize memory consumption, ensuring the model remains resource-efficient. These advancements collectively contribute to a more robust and scalable object tracking framework, capable of addressing diverse challenges in real-world applications.

2.1.2. Application of Object Tracking Technology

Object tracking technology finds extensive applications in areas such as intelligent security systems, autonomous driving, and drone navigation. In mobile scenarios, the demand for real-time performance is particularly high [2]. For instance, drones are required to perform dynamic obstacle avoidance at a frame rate of 30 FPS, ensuring safe and efficient navigation. Similarly, augmented reality (AR) devices depend on low-latency tracking to maintain the accurate overlay of virtual objects onto real-world environments. With a compact size of just 1.9MB, lightweight models can be seamlessly integrated into smartphones or edge devices, enabling advanced customer behavior analysis in smart retail environments. Furthermore, the integration of object tracking technology with federated learning has facilitated cross-camera collaborative tracking, enhancing the efficiency of urban management systems while simultaneously strengthening privacy protection in smart cities.

2.2. Overview of Lightweight Visual Model

2.2.1. The Characteristics and Advantages of Lightweight Visual Model

Lightweight visual models are designed to significantly reduce parameters and computational load while maintaining high levels of accuracy, offering several notable advantages. Firstly, these models excel in real-time performance, enabling rapid processing speeds, such as achieving detection rates of up to 150 frames per second on mobile devices. Secondly, they are optimized for energy efficiency, utilizing compatibility with ARM CPU or NPU architectures to minimize power consumption, which is particularly beneficial for battery-powered devices. Thirdly, they provide versatile deployment options, supporting various inference frameworks like TensorRT and NCNN, ensuring adaptability across diverse platforms [6]. For instance, architectures such as MobileNet achieve substantial reductions in computational complexity by employing techniques like channel reduction and width multiplier adjustments. Additionally, these models incorporate dynamic inference mechanisms that intelligently adjust the network's depth based on the computational capabilities of the device, thereby ensuring an optimal balance between performance and resource utilization.

2.2.2. Common Lightweight Visual Models

Mainstream lightweight models are widely recognized for their efficiency and adaptability in various applications, particularly in scenarios requiring optimized performance on resource-constrained devices [7].

MobileNet employs depthwise separable convolution to decouple spatial and channel features, significantly reducing computational complexity [1]. Its V3 iteration incorporates an advanced attention mechanism, enabling it to achieve an impressive 75.2% accuracy in the ImageNet classification task while maintaining a compact structure with only 5.4 million parameters.

LightTrack is a model specifically designed for tracking tasks, utilizing a neural architecture search (NAS) approach to optimize both accuracy and speed. By employing a multi-objective search strategy, it achieves a threefold increase in GPU inference speed compared to traditional models like SiamRPN++, making it highly efficient for real-time applications.

The nano tracker, a lightweight deep neural network solution, boasts a model size of just 1.9MB [8]. Despite its minimal size, it achieves a 68.5% success rate on the OTB100 dataset and supports real-time tracking capabilities on low-power devices, demonstrating its practical utility in constrained environments.

Baidu's PP-PicoDet is an ultra-lightweight detection model that integrates the ESNet backbone network with a CSP feature pyramid structure [9]. This design achieves a mean average precision (mAP) of 30.6% on the COCO dataset while maintaining a frame rate exceeding 150 frames per second on mobile devices, highlighting its exceptional balance of speed and accuracy.

These lightweight models, encompassing tasks such as classification, detection, and tracking, serve as the foundational technologies for enabling real-time object tracking on mobile devices. This is achieved through a combination of hardware and software optimizations, including neural processing unit (NPU) operator acceleration, ensuring seamless performance in diverse applications.

3. Challenges for Real-time Object Tracking on Mobile Devices

3.1. Computing Resources and Power Consumption Limitations

Mobile devices, including smartphones and drones, are typically equipped with low-power processors and limited memory bandwidth, which impose significant constraints on computing resources and energy consumption. These limitations present stringent requirements for target tracking algorithms. Traditional deep learning models, characterized by their extensive number of parameters, often exhibit high per-frame inference energy consumption, reaching up to 1-2 watts. This can result in severe device overheating and a noticeable reduction in battery life. Processing high-resolution video streams, such as 1080P, frequently necessitates the use of GPUs or NPUs. However, the heterogeneous computing architectures commonly employed in mobile hardware often suffer from insufficient memory read/write bandwidth, leading to computational bottlenecks. While model compression techniques can alleviate computational demands, excessive compression may compromise the model's feature representation capabilities. This is particularly evident in scenarios involving complex backgrounds or small targets, where accuracy tends to decline significantly. Consequently, deploying lightweight models requires a careful balance between accuracy and efficiency. For example, reducing the channel count of MobileNetV3 to 50% of its original configuration results in a 4.7% drop in detection accuracy, while energy consumption decreases by 62%. Achieving efficient inference under the constraints of limited CPU power, memory capacity, and battery life remains a critical challenge in the field of mobile object tracking, necessitating innovative solutions to optimize performance without sacrificing accuracy.

3.2. Real-time Requirements

The primary challenge in mobile object tracking is meeting the stringent requirement for real-time operation. In various application scenarios, such as unmanned aerial vehicle (UAV) obstacle avoidance or augmented reality (AR) interaction, maintaining a smooth visual experience is critical. This necessitates achieving at least 30 frames per second, with the total delay from initiation to completion controlled within 33 milliseconds. However, traditional algorithms, including correlation filtering and optical flow methods, often become computationally burdensome in complex environments. For instance, tracking objects in crowded scenes can result in processing times exceeding 80 milliseconds per frame, which significantly hampers real-time responsiveness.

Lightweight models aim to reduce computational demands through structural optimizations, yet they still encounter challenges such as prolonged feature matching times. These delays are often caused by variations in the size of dynamic objects and the presence of motion blur [10]. Additionally, mobile devices frequently perform multiple tasks simultaneously, such as camera preview and data transmission, which can lead to competition for computational resources. This competition may cause inconsistencies in the latency of the tracking thread, further complicating real-time performance.

To address these challenges, it is essential to incorporate hardware acceleration techniques, such as neural processing unit (NPU) operator optimization, to enhance computational efficiency. Furthermore, implementing advanced task scheduling mechanisms, such as priority queues, can help stabilize the tracking thread's performance [8]. Ensuring cross-platform compatibility adds another layer of complexity, as it requires managing the diverse hardware configurations of both iOS and Android devices. These measures collectively aim to maintain the real-time stability and reliability of mobile object tracking systems.

3.3. Complex Scene Adaptability

3.3.1. Varying Illumination

In the domain of real-time object tracking on mobile devices, variations in lighting conditions represent a common and challenging scenario that can substantially affect tracking accuracy. Fluctuations in ambient light cause noticeable changes in the external features of the tracked object. Under intense illumination, the colors of an object appear more vibrant, while the overall contrast within the image tends to decrease. On the other hand, in dimly lit environments, the target object becomes less distinguishable, and critical structural details may be obscured or lost entirely. These lighting dynamics introduce significant complexities in maintaining consistent tracking performance.

Changes in lighting conditions diminish the precision of feature extraction and complicate the process of matching the target object. Traditional object tracking algorithms are highly susceptible to variations in illumination; when the lighting environment shifts, issues such as target loss or tracking inaccuracies are likely to arise. To mitigate the adverse effects of light fluctuations, adaptive feature correction techniques tailored for illumination changes should be utilized [11]. Image preprocessing methods, including histogram equalization and color space conversion, can effectively adjust image brightness and contrast, thereby improving the stability and reliability of target features. Additionally, integrating multimodal information, such as combining visual and non-visual data, can further enhance tracking robustness in scenarios characterized by dynamic lighting conditions.

3.3.2. Object Occlusion

Blocking the target presents a significant challenge in real-time object tracking on mobile devices. In practical scenarios, the target being tracked is often partially or entirely obscured by other objects, leading to a loss of critical feature information. This loss complicates the ability of tracking algorithms to accurately determine the target's position and state. Such occlusions can disrupt the continuity of tracking and reduce the overall reliability of the system.

When a target becomes obscured, conventional tracking methods may fail to maintain accurate identification or continuity. Addressing this issue requires the development of advanced tracking state adjustment techniques. These methods can include predicting the target's trajectory, continuously estimating its position during periods of occlusion, and leveraging prior knowledge of the target's characteristics and its surrounding environment to rapidly reacquire it once the obstruction clears. Additionally, employing multi-target tracking technologies can enhance stability by monitoring the movements of nearby related objects. This approach allows for the estimation of the obscured target's location, ensuring more robust tracking performance in environments where occlusion is frequent. Such strategies are essential for improving the adaptability and reliability of tracking systems in dynamic and complex scenarios.

4. A Real-time Object Tracking Strategy for a Mobile Terminal Based on a Lightweight Visual Model

4.1. Lightweight Feature Extraction and Adaptive Optimization Strategy

4.1.1. Multi-scale Lightweight Feature Fusion Method

To address the challenges posed by variations in target size and background clutter, it is essential to enhance tracking accuracy by integrating features across multiple levels. This approach utilizes a lightweight backbone architecture to extract feature maps at varying depths, ensuring a balance between spatial detail and semantic richness. Shallow features, characterized by their detailed spatial information, are instrumental in pinpointing the precise location of the target. Conversely, deeper features, enriched with advanced semantic content, play a critical role in accurately identifying the target [12]. The lightweight feature pyramid module facilitates the fusion of multi-level features through techniques such as weighted summation and cross-level connections. This integration process introduces only a minimal computational overhead, enabling the model to effectively represent targets of diverse sizes. For mobile devices, the channel attention mechanism is employed to refine the integrated features by filtering out irrelevant background information and emphasizing the most significant attributes of the target. This strategy not only ensures a comprehensive representation of features but also enhances operational efficiency, providing optimal input for subsequent processes such as target matching and state updating. By leveraging these methods, the model achieves a robust balance between computational efficiency and tracking precision, making it highly suitable for real-time applications on mobile platforms.

4.1.2. Adjustment Based on Equipment Performance

On mobile devices, performance varies significantly, making it challenging to achieve both high accuracy on powerful devices and fast processing on less capable ones using a uniform feature extraction method. To address this, a mechanism is required that dynamically adjusts based on the performance of individual devices. Mobile phones can be categorized into distinct performance levels by evaluating critical parameters such as CPU and GPU computational capabilities, as well as memory bandwidth. For each performance tier, a predefined configuration of features is employed, which may involve modifications to the lightweight network's structure, including adjustments to its depth, width, or the resolution of input images. During the tracking process, the system continuously monitors the frame rate to ensure optimal functionality. If the frame rate drops below the established threshold, the system automatically transitions to a less demanding feature extraction mode to maintain real-time tracking. Conversely, when device resources are ample, the system intensifies feature extraction to achieve higher accuracy. This dynamic adjustment strategy fosters seamless collaboration between the algorithm and hardware, ensuring peak performance across a diverse range of mobile devices. By tailoring the computational load to the capabilities of each device, this approach maximizes efficiency and enhances the overall user experience.

4.2. Strategies for Enhancing Robustness in Complex Scenes

4.2.1. Calibration Method for Adaptive Illumination Change Feature

To address the issue of characteristic distortion caused by sudden changes in illumination, it is essential to develop an adaptive feature correction system [3]. This approach incorporates illumination-invariant features by transforming the original RGB image into the YCbCr or HSV color domain. This transformation leverages the separation of luminance (Y) and chrominance (CbCr) components, enhancing the model's ability to adapt to fluctuations in lighting conditions. Additionally, a lightweight illumination monitoring unit is introduced to evaluate parameters such as luminance distribution and frame contrast in real-time. This unit determines the extent of illumination changes and enables the feature extraction network to autonomously adjust its internal configurations. By doing so, the network can more effectively identify object edges and textures without over-relying on color information. Furthermore, a technique referred to as "short-term memory" is employed to integrate the stable target appearance retained prior to the illumination change with the current image features. This integration helps to mitigate the impact of abrupt feature alterations. By implementing these strategies, the system effectively reduces the interference caused by lighting variations, ensuring that the tracker operates reliably across diverse and challenging lighting environments. This comprehensive approach enhances robustness and adaptability, making it suitable for complex real-world scenarios.

4.2.2. Tracking State Correction Mechanism

To address the challenges of inaccurate tracking or lost tracking caused by occlusion and rapid movement, it is essential to develop a robust state adjustment system [4]. This system integrates motion prediction techniques with re-identification technology to ensure reliable tracking performance. When the target is partially obscured, predictive methods such as the Kalman filter or optical flow can be employed to estimate its trajectory based on prior state data. By reducing the reliance on appearance matching during these instances, the risk of incorrect updates due to insufficient feature information is minimized. In cases where the target is completely obscured or moves too rapidly for matching, a lightweight re-detection unit is activated. This unit predicts the target's center coordinates and deploys a small-scale detector to locate the target within a confined area. Once the target is successfully reacquired, the tracker’s state information is promptly updated, and the model is refined using the newly obtained appearance features. This iterative process of prediction and re-detection ensures adaptability to complex and dynamic environments, significantly enhancing tracking stability and reliability. By employing this mechanism, the system can effectively mitigate disruptions and maintain consistent performance in challenging scenarios.

5. Conclusion

The primary focus of this paper is that our mobile devices, such as mobile phones and tablets, encounter significant challenges, including insufficient computing power, slow response times, and complex surrounding environments. It systematically studies how to achieve real-time object tracking using a lightweight visual model. This research aims to resolve the contradiction between limited hardware performance and the demands of advanced tracking tasks, emphasizing the importance of optimizing computational efficiency while maintaining high accuracy. By addressing these challenges, the study provides a foundation for enhancing the usability of mobile devices in dynamic and resource-constrained scenarios.

Through in-depth analysis and rigorous demonstration, it reveals the three constituent elements of the core strategy: First, for feature extraction, it is necessary to leverage multi-scale lightweight feature fusion and dynamic adjustment mechanisms to significantly reduce computational costs while ensuring comprehensive information. Second, for model deployment, this paper integrates model pruning and incorporates mobile-optimized inference engines to achieve the ultimate streamlining of the tracking model in terms of size and speed, laying a solid foundation for real-time processing. Third, for complex real-world application scenarios, this paper proposes a novel calibration technique to address target occlusion and designs a fast motion tracking state correction mechanism, thereby significantly enhancing the system's robustness and reliability. These strategies collectively ensure that the proposed approach is not only computationally efficient but also adaptable to diverse and challenging environments, making it a practical solution for real-world applications.

This research combines a lightweight visual model with special optimization technology to create an efficient, reliable, and flexible real-time object tracking scheme for mobile devices. This scheme strikes a balance between accuracy and efficiency, making it more feasible for practical use in complex environments. Furthermore, it highlights the potential for future advancements in this field. With the enhancement of hardware performance and the continuous deepening of algorithm theory, achieving more precise and longer-lasting tracking while maintaining extremely low energy consumption remains a critical goal. Additionally, exploring deeper integration with semantic cognition and developing systems capable of understanding contextual information could open new avenues for innovation. These advancements would not only improve tracking performance but also expand the scope of applications, making such systems indispensable in areas like autonomous navigation, augmented reality, and intelligent surveillance. As such, this research lays a solid groundwork for addressing current limitations while inspiring further exploration into more sophisticated and sustainable tracking solutions.

References

1. S. Sun, N. Akhtar, H. Song, A. Mian, and M. Shah, "Deep affinity network for multiple object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 104–119, 2019.
2. W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, and T. K. Kim, "Multiple object tracking: A literature review," *Artificial Intelligence*, vol. 293, p. 103448, 2021.
3. B. Yan, H. Peng, K. Wu, D. Wang, J. Fu, and H. Lu, "Lighttrack: Finding lightweight neural networks for object tracking via one-shot architecture search," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15180–15189, 2021.
4. X. Weng, J. Wang, D. Held, and K. Kitani, "3D multi-object tracking: A baseline and new evaluation metrics," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 10359–10366, Oct. 2020.
5. G. Ciaparrone, F. L. Sánchez, S. Tabik, L. Troiano, R. Tagliaferri, and F. Herrera, "Deep learning in video multi-object tracking: A survey," *Neurocomputing*, vol. 381, pp. 61–88, 2020.
6. G. Chandan, A. Jain, and H. Jain, "Real time object detection and tracking using deep learning and OpenCV," in *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 1305–1308, Jul. 2018.
7. H. Fan, L. Lin, F. Yang, P. Chu, G. Deng, S. Yu, et al., "Lasot: A high-quality benchmark for large-scale single object tracking," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5374–5383, 2019.
8. Y. Cai, J. Liu, J. Tang, and G. Wu, "Robust object modeling for visual tracking," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9589–9600, 2023.
9. P. Mittal, "A comprehensive survey of deep learning-based lightweight object detection models for edge devices," *Artificial Intelligence Review*, vol. 57, no. 9, p. 242, 2024.
10. K. Hao, Z. Deng, B. Wang, Z. Jin, Z. Li, and X. Zhao, "Lightweight multiobject ship tracking algorithm based on trajectory association and improved YOLOv7tiny," *Expert Systems with Applications*, vol. 259, p. 125129, 2025.
11. C. Y. Tsai and Y. K. Su, "MobileNet-JDE: A lightweight multi-object tracking model for embedded systems," *Multimedia Tools and Applications*, vol. 81, no. 7, pp. 9915–9937, 2022.
12. S. B. J. Khan, P. Zhang, M. M. Kamal, and A. K. J. Saudagar, "LightTrack-ReID: A lightweight and occlusion-robust framework for multi-object tracking," *PloS One*, vol. 21, no. 3, p. e0342246, 2026.