International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 2, Mar-Apr 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

A Study on Object Tracking in Dynamic Scenes Using an LSTM-Attention Model

Sun Chenyang
Zhejiang Business Technology Institute, Ningbo, Zhejiang, 315012, China
<cysun@zbti.edu.cn>

----------**********----------

Abstract:

In dynamic scenes, object tracking faces challenges such as complex background interference and rapid object movement. To explore the application of object tracking technology in dynamic scenes, this paper utilises a visual model based on LSTM-Attention, aiming to improve the accuracy of object tracking in dynamic scenes. This model combines the memory capabilities of LSTM in processing time-series data with the advantages of the Attention mechanism in capturing the dynamic characteristics of objects in dynamic scenes. Building on this foundation, the model performs object tracking tasks in dynamic scenes. Its performance is compared with that of traditional object tracking algorithms, and the impact of various scene factors on tracking results is analysed to validate the advantages of the new model. Experimental data indicate that the model can accurately and stably track objects even when they are moving rapidly or partially occluded, maintaining an accuracy of 80.4% in complex dynamic scenes and demonstrating strong resistance to interference.

Keywords —**Dynamic scenes; Object tracking; LSTM-Attention**

----------**********----------

With rapid advances in the field of computer vision, object tracking—as a core branch of this discipline—is gradually finding applications in a number of key sectors, including intelligent transport systems, video surveillance and human-computer interaction. In dynamic and complex environments, the accuracy and robustness of object tracking have become crucial indicators of its performance. Such environments are often characterised by frequent occlusions, sudden changes in object appearance and unstable lighting conditions, posing significant challenges to traditional tracking methods.

Niu Sijie et al.[1],building upon the traditional KCF algorithm, extracted and fused CN features. By utilising complementary and symmetrical features to achieve multi-feature fusion, they effectively resolved the inaccuracies in target tracking caused by scale variations, thereby enhancing the algorithm's accuracy and success rate. Guo Chong et al.[2] innovatively proposed a convolutional hybrid attention mechanism, focusing on handling channel attention and spatial attention, which further improved the accuracy of target detection.

This paper proposes an object tracking technique for dynamic scenes based on the LSTM-Attention model, providing an in-depth analysis of the model's principles and advantages, and exploring its application effectiveness in complex dynamic environments. By effectively learning and memorising long-term dependencies within input

sequences[3], LSTM (Long Short-Term Memory) is capable of capturing the temporal motion characteristics of objects. Meanwhile, the Attention mechanism dynamically adjusts the attention weights for different regions[4], enabling the model to focus on the most critical information for the current tracking task when processing complex scenes. This enhances tracking accuracy and efficiency, thereby further strengthening the model's capabilities. The integration of LSTM and the Attention mechanism for target tracking in dynamic scenes resolves the issues faced by traditional methods—namely, the difficulty in handling targets' temporal motion characteristics and complex background information in complex dynamic scenes, which leads to low tracking accuracy and efficiency. This paper aims to provide an efficient and accurate technical solution for object tracking in dynamic scenes, thereby promoting the application and development of computer vision technology across a wider range of fields.

1. Data processing

1.1 Data Collection

This paper selects three representative dynamic scene datasets—UAV123, OTB50 and VOT2016—as the basis for model training and performance evaluation. These three datasets each possess distinct characteristics in terms of environmental types, object categories and motion patterns.

1.2 Data verification

During the data preparation phase, the first step is to download the compressed files for each dataset from the official website and ensure the integrity of all materials. Once the downloads are complete, a comprehensive data verification process is carried out. For the UAV123 and OTB50 datasets, the integrity of the video sequence files is verified in detail on a case-by-case basis; this involves checking whether each video file opens correctly, as well as verifying information such as file size and duration. By randomly playing back selected video clips, carefully observe whether there are any issues such as corrupted footage, stuttering, or missing segments, in order to further verify the integrity of the video content. At the same time, rigorously check the consistency between the annotation files and the video content, verifying each piece of annotated information line by line—including target location, class, and start and end times of tracking—to ensure that the annotation files correspond exactly and accurately to the recorded content, thereby avoiding the impact of annotation errors on the accuracy of model training and evaluation.

1.3 Data pre-processing

To enhance the efficiency and accuracy of image processing, a series of measures have been implemented to optimise each frame.

Image cropping. For each frame, precise cropping is performed on the original image based on the specific location of the target. This step not only removes a significant amount of redundant background area—thereby substantially reducing the computational load and improving processing efficiency—but also effectively minimises the interference caused by background noise on target detection. In complex image environments, background noise often obscures the target's feature information; precise cropping allows the model to focus its attention more intently on the target itself, creating favourable conditions for subsequent processing steps.

Image normalisation. Specifically, this involves normalising the image's pixel values to the range [0,1]. This standardisation significantly improves data processing efficiency during model training. Consistency and standardisation of data are crucial in model training. By normalising pixel values to a specific range, the complex adjustment processes caused by excessive variation in pixel values are avoided, enabling the model to perform computations in a more uniform manner when processing different images. This consistent processing approach enhances the model's ability to respond consistently to different input images; regardless of the distribution of the original pixel

values in the input image, the model can process them more stably, thereby improving overall recognition accuracy.

2. Model design

To construct the LSTM-Attention model, this paper has designed a temporal data processing module centred on an LSTM network; the internal structure of the LSTM is shown in Figure 1.

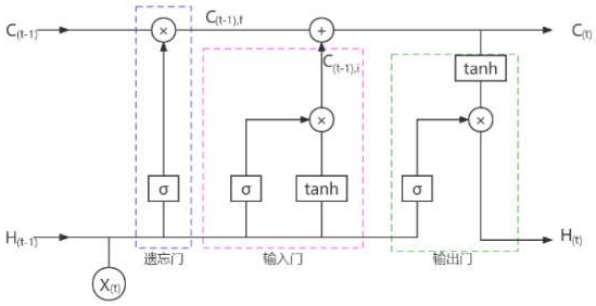

Fig 1 Diagram of the internal structure of an LSTM

2.1 An Overview of LSTM

In dynamic scenarios, the movement of targets exhibits complex time-series characteristics, and the primary purpose of the time-series processing module is precisely to efficiently handle long-range dependencies within time-series data. It acts like a precise detector, delving deep into the data to uncover hidden information and accurately capturing the dynamic characteristics of targets within the time series. In this way, the model gains a deeper understanding of changes in the target's position, velocity and other parameters at different points in time, thereby achieving a more accurate grasp of the target's trajectory. This is akin to endowing the model with 'keen eyesight', enabling it to track targets clearly amidst complex temporal information.

2.2 Attention mechanism

Relying solely on LSTM networks may not fully meet the model's requirements for capturing key information in complex scenarios; therefore, this paper introduces an attention mechanism, a schematic diagram of which is shown in Figure 2.

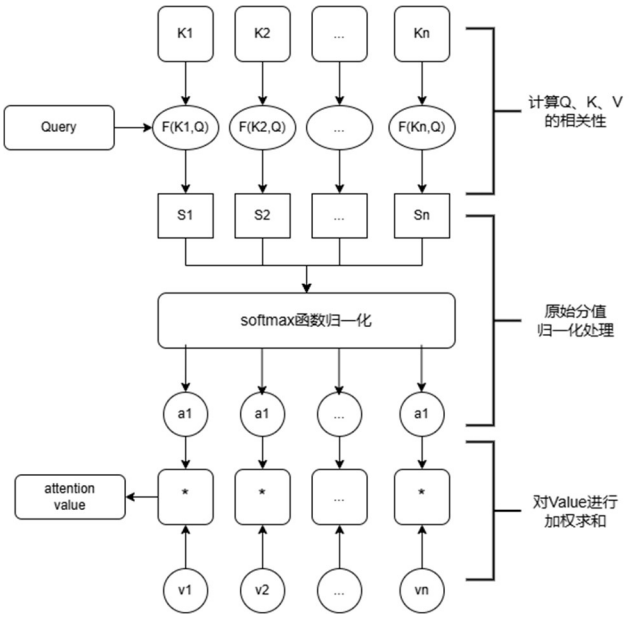

Fig 2: Schematic diagram illustrating the principle of the attention mechanism

3. Experimental Results and Analysis

3.1 Static background experiment

In a static background test environment, the LSTM-Attention model demonstrated exceptionally high accuracy, reaching 95.8%—a significant improvement over the 89.2% achieved by traditional algorithms—thus fully demonstrating the model's superiority in stable environments. This superiority is evident in the model's ability to accurately capture target features and achieve stable tracking, enabling precise target localisation against a static background with virtually no interference from other irrelevant factors.

3.2 Experiments on scenes with fast-moving targets

In challenging scenarios involving rapidly moving targets, further experimental results demonstrate that the model exhibits exceptional adaptability. Even under the complex conditions of rapid target movement, the model maintains a relatively high level of accuracy, reaching 87.3%, which is significantly higher than the 76.5% achieved by traditional algorithms. However, it

should be recognised that rapid movement does, to some extent, affect the model's tracking performance. In certain frames, some tracking errors are observed; this is primarily due to the fact that the target's position and shape change significantly within a short period of time as a result of rapid movement, thereby increasing the difficulty of the model's tracking and prediction tasks.

3.3 Target Occlusion Experiment

In the target occlusion experiments, the LSTM-Attention model continued to demonstrate strong performance in complex partial occlusion scenarios, achieving an accuracy of 82.6%, which is significantly higher than the 68.7% achieved by traditional algorithms. This is primarily attributable to the introduction of the attention mechanism, which effectively enhances the model's focus on key areas, enabling it to concentrate on the unoccluded critical parts even when the target is partially occluded, thereby mitigating the negative impact of occlusion. However, when the target is fully occluded, the model's performance drops markedly, with accuracy falling to 65.9%. This indicates that, in extreme occlusion scenarios, further research and refinement of the model's occlusion handling mechanisms are required to address the severe impact this has on target tracking, thereby further enhancing the model's robustness in complex environments.

4. Conclusion

This paper conducts an in-depth investigation into an LSTM-Attention-based object tracking model for dynamic scenes. Experimental results demonstrate that the model exhibits significant advantages in terms of accuracy when dealing with complex, dynamically changing scenes, and shows great potential in fields related to real-time tracking tasks, such as autonomous driving and intelligent surveillance. Its processing speed is capable of meeting the requirements of most real-time application scenarios, providing a valuable solution to object tracking problems in these fields. However, this study also highlights the model's limitations: the LSTM-Attention model has a high computational complexity. In resource-constrained environments, particularly when processing large-scale datasets or high-resolution images, it places high demands on hardware resources, which may increase deployment costs and hinder the model's widespread adoption. To address these issues, future research will focus on the following areas: firstly, further optimising the model architecture by adopting lightweight designs or introducing more efficient computational units to reduce computational costs; secondly, conducting research into multimodal fusion techniques to enhance the model's perception capabilities in complex scenarios through the integration of multi-source information.

ACKNOWLEDGMENT

Zhejiang Business Technology Institute, 2025 Institutional Research Grant Project: "Research on Multi-scale Dynamic Tracking Algorithms Based on Lightweight LSTM"（KYND202505）.

REFERENCES

[1] Niu Sijie, Wang Zhifeng, Zhu Jingjing. Target Tracking Based on Adaptive Scale Transformation and Feature Fusion [J]. Command, Control and Simulation, 2024, 46(04): 82–87.
[2] Guo Chong, Liu Sheng, Zhang Wenbo, et al. A Multi-Target Tracking Algorithm Based on a Convolutional Hybrid Attention Mechanism [J]. Control and Decision, 2024, 39(11):1-9.
[3] Liu Haodong. Research on Multi-Object Tracking Algorithms Based on Deep Learning [D]. Wuxi: Jiangnan University, 2023.
[4] Wu Yiwei. Research on Object Tracking Based on Attention Mechanisms [D]. Guangzhou: Guangzhou University, 2024.

