

Attention-Based Maneuver-Aware Tracking Network for Maneuvering Target Tracking

Jiahao Kang^{ID}, Graduate Student Member, IEEE, Haohao Ren^{ID}, Lin Zou^{ID}, Member, IEEE,
Jie Lin^{ID}, Member, IEEE, and Yun Zhou^{ID}

Abstract—Maneuvering target tracking is always a challenging problem due to the complexity of target motion state. Especially with highly maneuvering target, existing tracking algorithms struggle to swiftly and accurately respond to the sudden changes in target motion. In this letter, it is proposed for the first time that we should pay attention to the non-stationarity of maneuvering trajectory segments, and a self-attention-based maneuvering target tracking framework is developed. Specially, the proposed method first resorts to the maneuvering factors acquired from the mean and variance of trajectory segments to extract non-stationary maneuvering information, and then relies on the long-term dependence between observation segments to achieve maneuvering target tracking. Additionally, to enhance the robustness of the tracking model under the sudden change of motion patterns, a local feature embedding module is proposed to extract dynamically the local motion information of maneuvering target. Numerical experiments demonstrate the superiority of our proposed method over advanced deep learning-based maneuvering target tracking methods in enhancing modeling capabilities and improving robustness under abrupt changes in motion patterns.

Index Terms—Maneuvering target tracking, deep learning, attention mechanism, maneuvering factor.

I. INTRODUCTION

MANEUVERING target tracking is one of the most crucial tasks in various application fields, including airspace surveillance, ocean monitoring and autonomous navigation [1]. Due to the uncertainty of the target motion, it has always been a very challenging and highly concerned issue. Over the past few years, multi-model algorithms for modeling maneuvering target tracking as a mixture estimation problem [2] have been emerging [3], [4]. Among them, interactive multi-model (IMM) [5] and its variants [6], [7], [8] are generally considered to be effective solutions for solving maneuvering target tracking in the early days. Regrettably,

Received 13 February 2025; revised 6 May 2025; accepted 13 May 2025. Date of publication 19 May 2025; date of current version 15 July 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62201124 and Grant 42027805 for funding and in part by the Fundamental Research Funds for the Central Universities. The associate editor coordinating the review of this article and approving it for publication was Prof. Xuesong Wang. (*Corresponding author: Haohao Ren*)

Jiahao Kang, Haohao Ren, Lin Zou, and Yun Zhou are with the School of Information and Communication Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu 611731, China (e-mail: hao_ren@uestc.edu.cn).

Jie Lin is with the School of Aeronautics and Astronautics, Xihua University, Chengdu 610039, China (e-mail: xiangjianmchang@163.com).

Digital Object Identifier 10.1109/LSP.2025.3571402

IMM often encounters challenges such as model mismatches and significant time delays, particularly in complex application scenarios.

With the flourishing development of deep learning, a series of deep learning-based tracking methods have been proposed in succession. In earlier years, recurrent neural network (RNN) and its variants was applied to the maneuvering target tracking problem. In [9], Gao et al. presented a long short-term memory network (LSTM)-based tracking method to deal with the uncertainty of target motion. In [10], Jouaber et al. developed a neural network adapted Kalman filter integrating the merits of neural network and Kalman filter to achieve target tracking. Liu et al. proposed a bi-LSTM-based tracking method [11], which can quickly track maneuvering targets after training on massive off-line trajectory segment data. Nevertheless, these methods might suffer from the problem of tracking accuracy decline and even failure in the face of long observation segments due to the accumulation of errors and limitations in feature extraction capabilities.

Recently, Transformer [12] has been widely applied in maneuvering target tracking due to its excellent ability to model long-range dependencies. Zhao et al. proposed a transformer-based network (TBN) [13] to capture the long short-term dependencies of target states from a global perspective. Zhang et al. successively proposed Transformer-based tracking network TrTNet [14] and TrMTT [15] to address the state estimation problem for highly maneuverable target tracking tasks. Shen et al. presented two Transformer-based trackers [16] for smoothing and predicting target states. Unfortunately, existing Transformer-based tracking methods have overlooked the non-stationary and local maneuver information of trajectories. As a consequence, their tracking accuracy and robustness during the actual tracking phase are not satisfactory.

In response to the above mentioned problems, this letter proposes a novel tracking method named attention-based maneuver-aware tracking network, which is capable of extracting local motion features and learn maneuvering factors to achieve high-precision tracking of maneuvering targets. To capture local maneuver information, we develop a local feature embedding module that dynamically extracts discriminative features using local attention. Then, leveraging the mean and variance of trajectory segments, we calculated maneuvering factors to represent non-stationary maneuvers. This formed the basis of our self-attention-based maneuvering factor learning architecture, enabling robust tracking of maneuvering targets.

II. PROPOSED METHOD

A. Problem Formulation

Herein, we focus on the problem of state estimation of maneuvering target tracking in \mathbf{X} - \mathbf{Y} plane. In simple terms, how to estimate the true trajectory state $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k]^\top$ of the target from the observation data $\mathbf{O} = [\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_k]^\top$ acquired by the radar system. Let the true state of the target, i.e., $\mathbf{x}_i = [x, y]^\top$ be the real trajectory position in the two-dimensional scene, and the observation data, i.e., $\mathbf{o}_i = [r, \theta]^\top$ denotes the distance and azimuth observed by radar system, in which r and θ represent polar radius and polar angle, respectively. The proposed method intends to construct a novel mapping from the observed trajectory to the real trajectory.

B. Observation Pre-Process

Considering that measurement data processing needs to be completed in Cartesian coordinates, the observation data $\mathbf{o}_k = [r, \theta]^\top$ acquired from polar coordinates should be transformed to $\mathbf{z}_k = [x, y]^\top$ in the Cartesian coordinate system as below:

$$x = r \cdot \cos(\theta), y = r \cdot \sin(\theta) \quad (1)$$

It is worth mentioning that except for coordinate transformation, the proposed method no longer processes the observation data.

C. Feature Embedding Module

As pointed out in previous work [17], stationarity is essential for the predictability of time series data, and it is equally important in trajectory segment data tracking. As a consequence, in order to enhance the predictability of trajectory data, it is necessary to promote the stability of trajectory distribution. In view of this, the trajectory segment data $\mathbf{Z} = [z_1, z_2, \dots, z_k]^\top$ can first be normalized to $\mathbf{Z}' = [z'_1, z'_2, \dots, z'_k]^\top$, yielding:

$$\mu_{\mathbf{Z}} = \frac{1}{T} \sum_{i=1}^T z_i, \sigma_{\mathbf{Z}}^2 = \frac{1}{T} \sum_{i=1}^T (z_i - \mu_{\mathbf{Z}})^2 \quad (2)$$

$$z'_i = \frac{1}{\sigma_{\mathbf{Z}}} \odot (z_i - \mu_{\mathbf{Z}}) \quad (3)$$

where \odot is the element-wise product operation.

Afterwards, in order to learn maneuvering information for maneuvering targets, the first issue is how to mine discriminative features from trajectory segment data. For this matter, a local feature embedding module based on attention mechanism is developed, which strives to dynamically adjust the convolution kernel filter according to the trajectory segment data to extract discriminative features. Mathematically, the local dynamic convolution operations are as follows:

$$y = g \left(\mathbf{W}^\top (z') z' + \tilde{b}(z') \right) \quad (4)$$

$$\mathbf{W}(z') = \sum_{k=1}^K \pi_k(z') \mathbf{W}_k, \tilde{b}(z') = \sum_{k=1}^K \pi_k(z') \tilde{b}_k \quad (5)$$

where \mathbf{W}_k and b_k are the parameters to be learned for the k th convolution kernel, π_k is the weight of each convolution kernel, its calculation can refer to [18], and satisfies $0 \leq \pi_k(z') \leq 1$ and $\sum_{k=1}^K \pi_k(z') = 1$

D. Maneuvering Learning Mechanism

The non-stationary information within trajectory segments tends to be disregarded due to the corresponding normalization operation during feature embedding. However, such non-stationary information frequently stems from sudden changes in motion patterns, which is crucial for maneuvering target tracking. For that reason, we develop a maneuvering learning mechanism based on self-attention [19]. First, two maneuvering factors with non-stationary information are defined, which can be acquired by a shared MLP in terms of the original trajectory segments \mathbf{Z} and mean $\mu_{\mathbf{Z}}$ or variance $\sigma_{\mathbf{Z}}$ respectively, i.e.,

$$\log \varphi = MLP(\sigma_{\mathbf{Z}}, \mathbf{Z}) \quad (6)$$

$$\Phi = MLP(\mu_{\mathbf{Z}}, \mathbf{Z}) \quad (7)$$

Taking into account maneuvering factors with non-stationary information, the maneuvering learning mechanism can better focus on and fully learn the crucial aspects of maneuvering information. To extract the intrinsic correlation of data at each moment in the trajectory segment, a multi-head maneuvering attention layer is then devised, which is capable of mapping the input sequence to three feature map along different dimensions, i.e., the query (\mathbf{Q}) and key (\mathbf{K}), and then value(\mathbf{V}). Afterwards, the the intrinsic correlation between different positions in the trajectory sequences can be learned according to the following formula:

$$head_n(\mathbf{Q}', \mathbf{K}', \mathbf{V}', \varphi, \Phi) = \text{Softmax} \left(\frac{\varphi \mathbf{Q} \mathbf{K}'^\top + 1 \Phi^\top}{\sqrt{d_k}} \right) \mathbf{V}' \quad (8)$$

where $\mathbf{Q}' = (\mathbf{Q} - 1\mu_{\mathbf{Q}}^\top)/\sigma_{\mathbf{Z}}$. $1 \in \mathbb{R}^{S \times 1}$, $\mu_{\mathbf{Q}} \in \mathbb{R}^{d_k \times 1}$ and d_k denote an all-ones vector, the mean of \mathbf{Q} , and the key dimensionality, respectively. And so is the corresponding transformed \mathbf{K}' , \mathbf{V}' .

Subsequently, the multi-head attention mechanism concatenates maneuver information from each head across different representation subspace to learn diverse and complex features, i.e.,

$$\text{MultiHead}(\mathbf{Q}', \mathbf{K}', \mathbf{V}') = \text{Concat}(head_1, \dots, head_n) \quad (9)$$

E. Maneuvering Target Tracking

To achieve maneuvering target tracking, a fully connected layer is first designed for motion feature aggregation, and then de-normalization operation is conducted to map preliminary estimated states $\mathcal{H}(\mathbf{Z})$, denoted as $\tilde{\mathbf{X}}$, to the tracking results $\hat{\mathbf{X}}$ in the original space, as depicted in Fig. 1. Mathematically, the operations are as:

$$\hat{\mathbf{X}} = \mathcal{H}(\mathbf{Z}), \tilde{\mathbf{X}} = \sigma_{\mathbf{Z}} \odot (\hat{\mathbf{X}} + \mu_{\mathbf{Z}}) \quad (10)$$

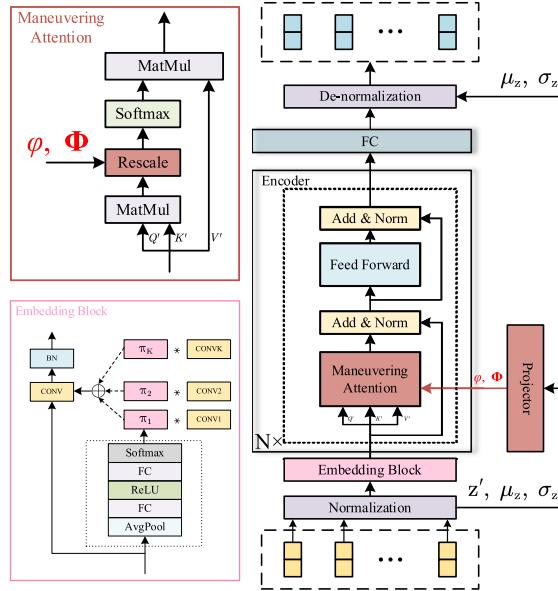


Fig. 1. The architecture of the proposed tracking network.

TABLE I
THE PARAMETERS OF LASTDATASET

Parameter	Value
Distance range	0.926km ~ 30.04km
Angle range	0° ~ 360°
Velocity range	0m/s ~ 340m/s
Turn rate	-10°/s ~ 10°/s
The standard deviation of acceleration noise	8m/s² ~ 13m/s²
The standard deviation of azimuth noise	0.401° ~ 0.516°
The standard deviation of distance noise	8m ~ 13m

III. EXPERIMENTS AND ANALYSIS

A. Dataset Introduction

To validate the effectiveness of the proposed tracking method, Lastdataset [11] dataset is leveraged, which consists of different observation trajectories from many different maneuvering targets in the X-Y coordinate based on the state space model (SSM). Table I lists the parameters of Lastdataset. Note that the proposed method only exploits the positional information of the target. The state transition equation and observation equation based on the SSM are as below:

$$\begin{cases} \mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + n \\ \mathbf{o}_t = h(\mathbf{x}_t) + m \end{cases} \quad (11)$$

where \mathbf{F} is state transition matrix. In the Lastdataset, two motion modes are considered, i.e., constant velocity (CV) and constant turn (CT). The transition matrix of CV and CT is defined as:

$$\mathbf{F}_{CV} = \begin{bmatrix} 1 & 0 & \tau & 0 \\ 0 & 1 & 0 & \tau \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (12)$$

$$\mathbf{F}_{CT} = \begin{bmatrix} 1 & 0 & \frac{\sin(\omega\tau)}{\omega} & \frac{\cos(\omega\tau)-1}{\omega} \\ 0 & 1 & \frac{1-\cos(\omega\tau)}{\omega} & \frac{\sin(\omega\tau)}{\omega} \\ 0 & 0 & \cos(\omega\tau) & -\sin(\omega\tau) \\ 0 & 0 & \sin(\omega\tau) & \cos(\omega\tau) \end{bmatrix} \quad (13)$$

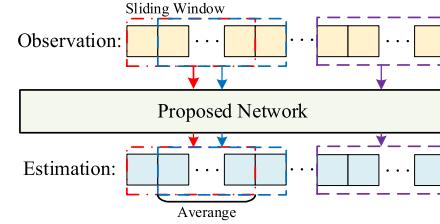


Fig. 2. Sliding window operation in long trajectory segment.

Where $\tau = 1s$ and ω represent sampling interval and turning rate. n and m denote transition noise and observed noise, respectively. Among them, the observation noise $m = [m_r, m_\theta]^\top$ consists of distance noise m_r and azimuth noise m_θ , where $m_r \sim \mathcal{N}(0, \sigma_r^2)$ and $m_\theta \sim \mathcal{N}(0, \sigma_\theta^2)$. Transition noise $n = [n_{p,k}, n_{p,k}, n_{v,k}, n_{v,k}]^\top$ can be calculated as below [20]:

$$[n_{p,k}, n_{p,k}, n_{v,k}, n_{v,k}]^\top = \begin{bmatrix} \frac{\tau^2}{2} & \frac{\tau^2}{2} & 0 & 0 \\ 0 & 0 & \tau & \tau \end{bmatrix} \cdot [\alpha_k \quad \alpha_k]^\top \quad (14)$$

where $\alpha_k \sim \mathcal{N}(0, \sigma_a^2)$. In addition, $h(\cdot)$ denotes the nonlinear transformation from $\mathbf{X}_t = [x, y]^\top$ to observation data $\mathbf{O}_t = [r, \theta]_t^\top$. \mathbf{O}_t is defined as

$$\mathbf{O}_t = \begin{bmatrix} \sqrt{x_t^2 + y_t^2} \\ \arctan(y_t/x_t) \end{bmatrix} + \begin{bmatrix} m_r \\ m_\theta \end{bmatrix} \quad (15)$$

B. Tracking Experiment Settings

A sliding window with length $L = 10$ and step size $K = 1$ is used to process long observation track segments, as shown in Fig. 2. By doing so, the latest measurement and the state values from the previous nine steps form the input trajectory segment of the tracking model. Moreover, The overlapping region is averaged to obtain the final state estimation.

In terms of the proposed tracking model, the depth of model is set to $N=3$. The embedding dimension d_{model} is set to 128. For bigger feature space, we employ 8 heads and $d_{ff} = 4d_{model}$ in the multi-head attention layer. The Adam optimizer [21] with a learning rate of 0.005 is adopted for network training. All experiments are performed on a server with NVIDIA GeForce RTX 3090 Ti GPU.

In the following experiments, the root-mean-squared error (RMSE) is leveraged to measure the difference between the model predicted trajectory and the real trajectory. RMSE is defined as follows:

$$Loss = \sqrt{\frac{1}{k} \sum_{i=1}^k (\hat{\mathbf{x}}_i - \mathbf{x}_i)^2} \quad (16)$$

C. Comprehensive Evaluation

In order to the effectiveness and superiority of the proposed tracking method, a complex maneuvering trajectory segment for 70 seconds named T1 is first simulated in this section. Let the initial position of the trajectory be [1000 m, 3000 m, 60 m/s, 20 m/s]. The turning rate and movement pattern for each segment trajectory are shown in Table II. Furthermore, the standard deviations of acceleration, azimuth,

TABLE II
DETAILED INFORMATION OF MANEUVERING TRAJECTORY

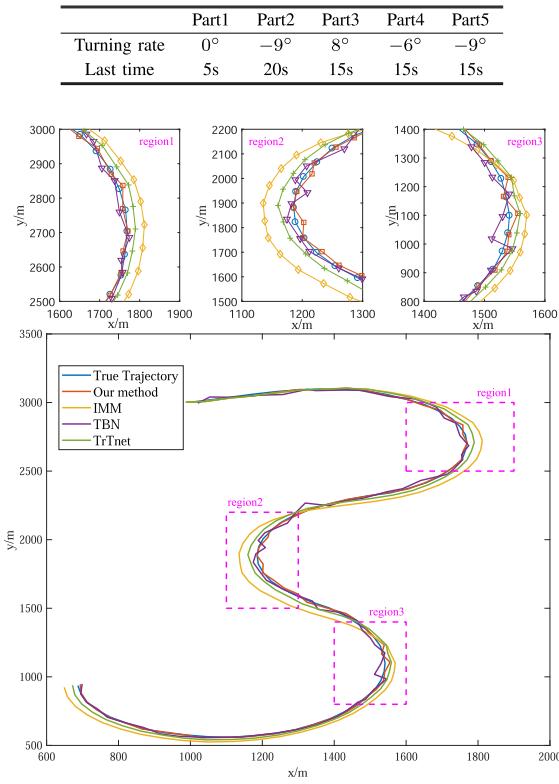


Fig. 3. Comparison of maneuvering target tracking results for each tracking method.

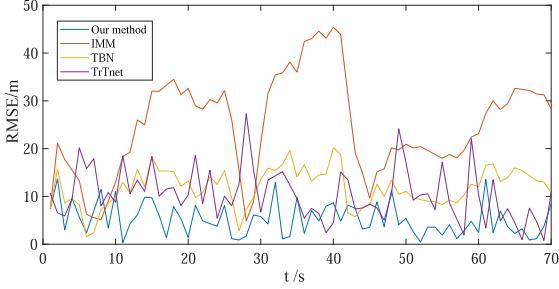


Fig. 4. Comparison of RMSE for each tracking method.

and distance noise are set to 5 m/s^2 , 0.25° , and 8 m , respectively. Several state-of-the-art maneuvering target tracking methods including IMM [22], TrTnet [14], and TBN [13] are employed as competitors in the following simulation experiments. The visualized experimental results are depicted in Fig. 3, in which the three sub-pictures are local magnification results of three regions. As can be observed from Fig. 3, the proposed tracking method is always superior to all competitors. In particular, one can see from the three sub-figures in Fig. 3 that the robustness of the proposed tracking method is best when the target is highly maneuverable.

Fig. 4 plots the the RMSE result for each tracking method. Notably, IMM has the worst tracking accuracy among all tracking methods, especially when the target is in high maneuvering, it

TABLE III
THE MEAN AND DEVIATION OF THE POSITIONS FOR EACH TRACKING METHOD (m)

	IMM		TBN		TrTnet		Our method	
	mean	Dev	mean	Dev	mean	Dev	mean	Dev
Part1	15.02	37.22	9.91	4.41	10.58	5.52	7.78	4.11
Part2	23.62	18.58	11.66	4.30	12.17	3.97	5.82	3.28
Part3	31.47	13.23	13.61	4.59	10.76	6.16	5.07	3.61
Part4	20.52	12.12	9.98	4.17	11.39	2.69	4.86	2.92
Part5	37.02	12.05	13.06	2.55	9.13	3.04	4.09	3.15

TABLE IV
ROBUSTNESS ANALYSIS FOR EACH TRACKING METHOD (m)

Method	IMM	TBN+CM	TrTnet	Our method
Avg-RMSE	46.32	22.54	24.52	13.01

shows significant tracking deviation. In additin, one can see from Fig. 4 that the proposed method outperforms both transformer-based tracking methods and IMM in segments with high turn rates. This is attributed to the fact that the proposed method fully considers and mines the maneuvering information in the trajectory data, while the tracking performance of competitors is restrained due to their failure to consider this. In order to analyze the tracking accuracy of the proposed method in statistical significance, 50 Monte Carlo experiments are conducted. The mean and deviation (Dev) of the RMSE for the positions of each method are shown in Table III. It can be seen from Table III that the proposed tracking method is statistically far better than each competitor in terms of RMSE.

To assess the generalization ability and robustness of the proposed tracking method, we increase the measurement amplitude noise standard deviation to 30 m to carry out the evaluation experiments. The average RMSE (Avg-RMSE) of target position estimates over the entire trajectory for each method are calculated in this experiment. The experimental result of each method is shown in Table IV. One can see from Table IV that the RMSE of all tracking methods has increased to varying degrees compared with the experimental results in Table 4, but the Avg-RMSE of the proposed tracking method is much smaller than that of each competitor.

IV. CONCLUSION

In this letter, a novel tracking framework-based deep learning named attention-based maneuver-aware tracking network is proposed to achieve high-precision maneuvering target tracking. The innovations of this letter an be summarized in two aspects. On one hand, the proposed method is the first to extract and utilize non-stationary maneuvering information from moving trajectories for maneuvering target tracking tasks. On the other hand, the proposed method integrates local maneuver information extraction and self-attention-based tracking architecture to enhance the robustness and representation capabilities of the tracking model. Simulation experiments illustrate that the proposed tracking method is competitive. In future work, we will integrate the track management algorithm to enhance the proposed algorithm, enabling it to achieve excellent performance in complex multi-target scenarios.

REFERENCES

- [1] J. Lim, H.-S. Kim, and H.-M. Park, "Interactive-multiple-model algorithm based on minimax particle filtering," *IEEE Signal Process. Lett.*, vol. 27, pp. 36–40, 2020.
- [2] L. Gao, M. Li, and Z. Jing, "Multiple maneuvering target tracking based on hierarchical dirichlet process and hidden Markov model," *Signal Process.*, vol. 217, 2024, Art. no. 109344.
- [3] T. R. Rice and A. T. Alouani, "Multiple-model filtering," *Acquisition Tracking Pointing XII*, vol. 3365, pp. 100–112, 1998.
- [4] S. McGinnity and G. W. Irwin, "Multiple model bootstrap filter for maneuvering target tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 3, pp. 1006–1012, Jul. 2000.
- [5] A. Munir and D. P. Atherton, "Adaptive interacting multiple model algorithm for tracking a manoeuvring target," *IEE Proc.-Radar Sonar Navigation*, vol. 142, no. 1, pp. 11–17, 1995.
- [6] U. Hammes and A. M. Zoubir, "Robust MT tracking based on M-estimation and interacting multiple model algorithm," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3398–3409, Jul. 2011.
- [7] R. Visina, Y. Bar-Shalom, and P. Willett, "Multiple-model estimators for tracking sharply maneuvering ground targets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 3, pp. 1404–1414, Jun. 2018.
- [8] Y. Di et al., "A maneuvering target tracking based on fastimm-extended viterbi algorithm," *Neural Comput. Appl.*, vol. 37, pp. 7925–7934, 2025.
- [9] C. Gao, J. Yan, S. Zhou, P. K. Varshney, and H. Liu, "Long short-term memory-based deep recurrent neural networks for target tracking," *Inf. Sci.*, vol. 502, pp. 279–296, 2019.
- [10] S. Jouaber, S. Bonnabel, S. Velasco-Forero, and M. Pilte, "NNAKF: A neural network adapted Kalman filter for target tracking," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2021, pp. 4075–4079.
- [11] J. Liu, Z. Wang, and M. Xu, "DeepMTT: A deep learning maneuvering target-tracking algorithm based on bidirectional LSTM network," *Inf. Fusion*, vol. 53, pp. 289–304, 2020.
- [12] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, vol. 30.
- [13] G. Zhao, Z. Wang, Y. Huang, H. Zhang, and X. Ma, "Transformer-based maneuvering target tracking," *Sensors*, vol. 22, no. 21, 2022, Art. no. 8482.
- [14] Y. Zhang, G. Li, X.-P. Zhang, and Y. He, "Transformer-based tracking network for maneuvering targets," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2023, pp. 1–5.
- [15] Y. Zhang, G. Li, X.-P. Zhang, and Y. He, "A deep learning model based on transformer structure for radar tracking of maneuvering targets," *Inf. Fusion*, vol. 103, 2024, Art. no. 102120.
- [16] L. Shen, H. Su, Z. Li, C. Jia, and R. Yang, "Self attention-based transformer for nonlinear maneuvering target tracking," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5109013.
- [17] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. Melbourne, VIC, Australia: OTexts, 2018.
- [18] Y. Chen, X. Dai, M. Liu, D. Chen, L. Yuan, and Z. Liu, "Dynamic convolution: Attention over convolution kernels," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11030–11039.
- [19] Y. Liu, H. Wu, J. Wang, and M. Long, "Non-stationary transformers: Exploring the stationarity in time series forecasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, vol. 35, pp. 9881–9893.
- [20] J. Liu, Z. Wang, and M. Xu, "A Kalman estimation based rao-blackwellized particle filtering for radar tracking," *IEEE Access*, vol. 5, pp. 8162–8174, 2017.
- [21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, San Diego, CA, USA, May 2015.
- [22] D. Magill, "Optimal adaptive estimation of sampled stochastic processes," *IEEE Trans. Autom. Control*, vol. AC-10, no. 4, pp. 434–439, Oct. 1965.