Extended Object Tracking: A DNN-aided approach

Runhe Gan, Wei Xia*†
School of Information and Communication Engineering,
University of Electronic Science and Technology of China, Chengdu 611731, China
rhg@std.uestc.edu.cn, wx@uestc.edu.cn
†School of Computer Science and Technology (School of Cyberspace Security), Xinjiang University, Ürümqi 830046, China

***Abstract*—In this work, we propose a lightweight and interpretable hybrid-driven extended object tracking (EOT) method that integrates deep neural networks with the conventional random matrix-based EOT algorithm based on a state space model. Two main issues in current methods are addressed herein, namely, performance degradation of model-based algorithms with partially unknown parameters of the state space model, and the high computational complexity and lack of interpretability of deep-neural-network-based approaches. We develop a positive definiteness and symmetry preservation scheme. Both the simulation and experimental results generated from the pre-processed synthetic marine X-band radar measurements demonstrate the superior performance of our method.**

I. Introduction

With the growing use of high-resolution radar sensors, such as X-band marine radar widely applied for vessel tracking [1], extended object tracking (EOT) has gained increasing attention [2], owing to its capability of jointly estimating both the kinematic and extension states of an object [3]. As a result, EOT plays a vital role in autonomous driving [4], maritime surveillance [5], and other fields.

In recent years, among the celebrated model-based (MB) inference EOT algorithms such as random hypersurfaces [6], particle filters [7], and random matrix (RM) [3], [8]–[11], RM-based algorithms have emerged as a representative, computationally efficient and interpretable [10], [11] approach. In RM-based methods, the extended object (EO) is typically modeled as an ellipse or a combination of ellipses, with the extension state modeled by a symmetric positive-definite (SPD) matrix governed by an Inverse-Wishart distribution. However, as experimentally validated hereinafter, the tracking performance of these MB algorithms would generally degrade in the scenarios that the state-space (SS) model parameters are imprecise, or even partially unknown, as the MB algorithms rely on the accurate SS model parameters.

The efficacy of Deep neural network (DNN)-based methods have also been validated for EOT [12]. Incorporating the decoupling technique, the DNN-based approaches could track the states of the kinematics and the shape of the EO with Kalman filters (KFs) and CNNs, respectively [13], [14]. Though these DNN-based methods could achieve superior performance in challenging real-world scenarios [14], they lack interpretability and are generally computationally expensive.

†This work was supported in part by the National Natural Science Foundation of China under Grants 62471107, 62461053. (*Corresponding author: W. Xia.*)

Considering the pros and cons of the MB and deep neural network (DNN)-based methods, hybrid-driven (HD) approaches that combine domain knowledge with data-driven learning have recently gained increasing attention [15]. Specifically, the structure-oriented DNN-aided inference methods [16], [17] could effectively track point targets by embeding DNN modules into MB algorithms to estimate critical statistics. Moreover, gated recurrent units (GRUs) have been demonstrated to be effective in learning temporal dependencies of statistics such as Kalman gains [16], while split architectures that decouple process-noise-related and measurement-noise-related estimations could enhance the robustness of point target tracking with partially unknown parameters [17]. Nevertheless, developing efficient hybrid-driven methods for EOT remains an open problem.

We propose herein a lightweight and highly interpretable DNN-aided EOT framework, which is robust in the tracking scenarios with partially unknown parameters. Our approach integrates DNNs into the conventional RM-EOT algorithm [9], leveraging the principles of structure-oriented DNN-aided inference [15]. Further, we develop a computationally efficient positive definiteness and symmetry preservation scheme to estimate the statistics obtained via the DNNs. Simulation results with the pre-processed synthetic marine X-band radar measurements demonstrate the superiority of the proposed method with partially unknown parameters.

II. Problem Formulation

We consider tracking a single EO, whose shape is modeled as an ellipse in the $d = 2$ space [3], [8], [9]. The EO's comprehensive state at time step k includes its kinematic state $\boldsymbol{x}_k \triangleq [\boldsymbol{p}_k^{\mathrm{T}}, \boldsymbol{v}_k^{\mathrm{T}}]^{\mathrm{T}}$ with $\boldsymbol{p}_k$ and $\boldsymbol{v}_k$ denoting the position and velocity, respectively, as well as its extension state $\boldsymbol{X}_k \triangleq \boldsymbol{\Omega}(\theta)\mathrm{diag}(l_1^2, l_2^2)\boldsymbol{\Omega}(\theta)^{\mathrm{T}} \in \mathbb{R}^{d\times d}$, where l_1 and l_2 are the semi-axes of the ellipse, and $\boldsymbol{\Omega}(\theta) = \begin{bmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{bmatrix}$ is the rotation matrix parameterized by orientation angle θ [9].

Consider the following dynamic model of the EO kinematic state [3], [8], [9],

$$\boldsymbol{x}_k = \tilde{\boldsymbol{F}}_k \boldsymbol{x}_{k-1} + \boldsymbol{w}_{k-1}, \quad \boldsymbol{w}_k \sim \mathcal{N}(\boldsymbol{0}_{2d}, \boldsymbol{D}_k \otimes \boldsymbol{X}_k), \quad (1)$$

with the covariance matrix $\boldsymbol{D}_k \in \mathbb{R}^{d\times d}$ of the corresponding Gaussian process noise in the one-dimensional (1-d) model [9]. Following [3], [9], we assume that $\tilde{\boldsymbol{F}}_k \triangleq \boldsymbol{F}_k \otimes \boldsymbol{I}_d$, with $\boldsymbol{F}_k$ denoting the 1-d state transition matrix, for the sake

979-8-3315-7206-8/25/$31.00

of simplicity in deriving the MB inference process of the proposed HD-EOT.

The evolution of the EO extension follows a Wishart evolution model [9]

$$p(\boldsymbol{X}_k \mid \boldsymbol{X}_{k-1}) = \mathcal{W}(\boldsymbol{X}_k; \delta_k, \boldsymbol{A}_k \boldsymbol{X}_{k-1} \boldsymbol{A}_k^{\mathrm{T}}), \tag{2}$$

with degrees of freedom $\delta_k > d-1$ and the invertible extension evolution matrix $\boldsymbol{A}_k \in \mathbb{R}^{d\times d}$ [9].

The original measurements of the EO are acquired by a marine X-band radar, and then pre-processed utilizing a Constant False Alarm Rate (CFAR) detector to produce polar detections (range and azimuth), which are subsequently transformed into Cartesian coordinates [18]. Each resulting measurement $\boldsymbol{z}_k^r$ could be modeled as

$$\boldsymbol{z}_k^r = \tilde{\boldsymbol{H}}_k \boldsymbol{x}_k + \boldsymbol{v}_k^r, \quad \boldsymbol{v}_k^r \sim \mathcal{N}(\boldsymbol{0}_d, \boldsymbol{B}_k \boldsymbol{X}_k \boldsymbol{B}_k^{\mathrm{T}}), \tag{3}$$

with the invertible matrix

$$\boldsymbol{B}_k = (\lambda \boldsymbol{X}_k + \boldsymbol{R}_k)^{1/2} \boldsymbol{X}_k^{-1/2}, \tag{4}$$

describing the observation distortion of the extension [9], [10], [19], with λ denoting a scaling factor for the spread contribution of the EO extension $\boldsymbol{X}_k$, and $\boldsymbol{R}_k$ the true measurement noise covariance matrix [8]. Following [3], [9], we assume that the measurement matrix $\tilde{\boldsymbol{H}}_k \triangleq \boldsymbol{H}_k \otimes \boldsymbol{I}_d$, with $\boldsymbol{H}_k$ denoting the 1-d measurement matrix.

For brevity, let $\hat{\boldsymbol{x}}_{k|j}$ and $\hat{\boldsymbol{X}}_{k|j}$ denote the instantaneous estimates of the kinematic state and the extension state at each time step $k \geq j$, respectively, utilizing the measurements up to time step j. Let $\boldsymbol{\Sigma}_{k|j} \triangleq \mathbb{E}\left\{(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|j})(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|j})^{\mathrm{T}}\right\}$ denote the covariance matrix of the kinematic state estimation error. Following [3], [9], we assume that $\boldsymbol{\Sigma}_{k|k-1} \triangleq \boldsymbol{P}_{k|k-1} \otimes \boldsymbol{X}_k$ and $\boldsymbol{\Sigma}_{k|k} \triangleq \boldsymbol{P}_{k|k} \otimes \boldsymbol{X}_k$, with $\boldsymbol{P}_{k|k-1}$ and $\boldsymbol{P}_{k|k}$ denoting the predicted and updated 1-d kinematic error covariance matrix, respectively.

The comprehensive state of the EO could be predicted in light of (1) and (2) in the prediction step. Whereas in the update step, in light of the Bayes' formula, the conditional probability density function (pdf) $p\left(\boldsymbol{x}_k, \boldsymbol{X}_k \mid \mathbb{Z}^k\right)$ is given by [9]

$$\begin{aligned} &p\left(\boldsymbol{x}_k, \boldsymbol{X}_k \mid \mathbb{Z}^k\right) \propto p\left(\mathbb{Z}_k \mid \boldsymbol{x}_k, \boldsymbol{X}_k, \mathbb{Z}^{k-1}\right) p\left(\boldsymbol{x}_k, \boldsymbol{X}_k \mid \mathbb{Z}^{k-1}\right) \\ &= \mathcal{N}\left(\bar{\boldsymbol{z}}_k; \boldsymbol{H}_k \boldsymbol{x}_k, \boldsymbol{B}_k \boldsymbol{X}_k \boldsymbol{B}_k^{\mathrm{T}} / n_k\right) \mathcal{W}\left(\bar{\boldsymbol{Z}}_k; n_k - 1, \boldsymbol{B}_k \boldsymbol{X}_k \boldsymbol{B}_k^{\mathrm{T}}\right) \\ &\cdot \mathcal{N}\left(\boldsymbol{x}_k; \hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1} \otimes \boldsymbol{X}_k\right) \mathcal{IW}\left(\boldsymbol{X}_k; \hat{v}_{k|k-1}, \hat{\boldsymbol{V}}_{k|k-1}\right), \end{aligned} \tag{5}$$

where the centroid measurement $\bar{\boldsymbol{z}}_k$ and the corresponding scattering matrix $\bar{\boldsymbol{Z}}_k$ are respectively given by [3]

$$\bar{\boldsymbol{z}}_k \triangleq \frac{1}{n_k} \sum\nolimits_{j=1}^{n_k} \boldsymbol{z}_k^j, \tag{6}$$

$$\bar{\boldsymbol{Z}}_k \triangleq \sum\nolimits_{j=1}^{n_k} \left(\boldsymbol{z}_k^j - \boldsymbol{z}_k\right) \left(\boldsymbol{z}_k^j - \boldsymbol{z}_k\right)^{\mathrm{T}}. \tag{7}$$

Based on this, a closed-form recursive algorithm, termed the RM-EOT could be derived, with specific equations presented hereinafter.

We consider the partially-unknown scenario where the parameters of the SS model are partially unknown. Specifically, we have the following assumptions

- A1 The measurement matrix $\tilde{\boldsymbol{H}}_k$ is assumed to be accurate.
- A2 The matrices $\tilde{\boldsymbol{F}}_k$ and $\boldsymbol{A}_k$ approximately describe EO dynamics, with unknown $\boldsymbol{D}_k$ and δ_k.
- A3 The observation distortion $\boldsymbol{B}_k$ is given by (4), with unknown $\boldsymbol{R}_k$.

In light of Assumptions A1, A2 and A3, we herein design a computationally efficient and robust DNN-aided method to track an EO with partially unknown state space model parameters, leveraging the hybrid inference methodology combining DNNs and MB algorithms.

III. DNN-AIDED HD-EOT METHOD

We now develop the DNN-aided HD-EOT method, which is interpretable. At each time step k, the HD-EOT follows a prediction-update structure akin to that of the RM-EOT. We use a dedicated neural network module, EOTNet, to estimate the statistics that may deteriorate due to partially unknown parameters.

A. Abstract Level Architecture

We first propose a positive definiteness and symmetry preservation scheme as an indispensable ingredient for the HD-EOT such that we could retain the SPD structure of the predicted 1-d state error covariance matrix $\boldsymbol{P}_{k|k-1}$, and ensure that the 1-d innovation covariance matrix [9]

$$S_k \triangleq \boldsymbol{H}_k \boldsymbol{P}_{k|k-1} \boldsymbol{H}_k^{\mathrm{T}} + \frac{|\boldsymbol{B}_k|}{n_k} \tag{8}$$

is positive. Specifically, we have $\boldsymbol{P}_{k|k-1} = \boldsymbol{L}_{p,k} \boldsymbol{L}_{p,k}^{\mathrm{T}}$ and $S_k = L_{s,k} L_{s,k}^{\mathrm{T}}$, with the following constraint

- C1 $\boldsymbol{L}_{p,k}$ is a lower triangular matrix with positive diagonal elements, and $L_{s,k}$ is a positive scalar.

Notice that both square root matrices in $\boldsymbol{B}_k$ (4) could be represented in the form of their Cholesky decomposition. Correspondingly, we have

- C2 $\boldsymbol{B}_k$ is a lower triangular matrix with positive diagonal elements.

Consequently, the task of the EOTNet becomes estimating $\boldsymbol{L}_{p,k}$, $L_{s,k}$, and $\boldsymbol{B}_k$ instead of directly estimating $\boldsymbol{P}_{k|k-1}$, S_k.

We then revamp the prediction of the conventional framework of the RM-EOT in the absence of the prior knowledge of the 1-d process noise covariance matrix $\boldsymbol{D}_k$. We utilize the EOTNet to predict $\boldsymbol{L}_{p,k}$, with the inference logic resembling

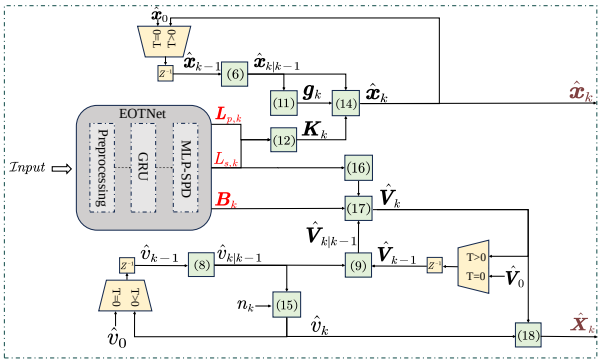


Fig. 1: Abstract Level architecture of the HD-EOT.

that of the RM-EOT [9]. Specifically, at each time step k,

$$\hat{\boldsymbol{x}}_{k|k-1} = (\boldsymbol{F}_k \otimes \boldsymbol{I}_d)\,\hat{\boldsymbol{x}}_{k-1}, \tag{9}$$

$$\lambda_{k-1} = \hat{v}_{k-1} - 2d - 2, \tag{10}$$

$$\hat{v}_{k|k-1} = \frac{2\delta_k\,(\lambda_{k-1}+1)\,(\lambda_{k-1}-1)\,(\lambda_{k-1}-2)}{\lambda_{k-1}^2\,(\lambda_{k-1}+\delta_k)} + 2d + 4, \tag{11}$$

$$\hat{\boldsymbol{V}}_{k|k-1} = \frac{\delta_k}{\lambda_{k-1}}\left(\hat{v}_{k|k-1} - 2d - 2\right)\boldsymbol{A}_k \hat{\boldsymbol{V}}_{k-1}\boldsymbol{A}_k^T, \tag{12}$$

$$\hat{\boldsymbol{X}}_{k|k-1} = \hat{\boldsymbol{V}}_{k|k-1} / \left(\hat{v}_{k|k-1} - 2d - 2\right). \tag{13}$$

We further develop an alternative update step in place of the counterpart in the RM-EOT, under Assumptions A2 and A3. $\boldsymbol{B}_k$ and $L_{s,k}$ would thus be obtained through the EOTNet. Specifically, akin to the square-root filtering strategy [20] and the inference logic of the RM-EOT, at each time step k,

$$\boldsymbol{g}_k \triangleq \bar{\boldsymbol{z}}_k - (\boldsymbol{H}_k \otimes \boldsymbol{I}_d)\,\hat{\boldsymbol{x}}_{k|k-1}, \tag{14}$$

$$\boldsymbol{K}_k = \boldsymbol{L}_{p,k}\boldsymbol{L}_{p,k}^{\mathrm{T}}\boldsymbol{H}_k^{\mathrm{T}} L_{s,k}^{-\mathrm{T}} L_{s,k}^{-1}, \tag{15}$$

$$\boldsymbol{L}_{e,k} = \boldsymbol{L}_{p,k} - \boldsymbol{K}_k L_{s,k}\left(L_{s,k} + \frac{\sqrt{|\boldsymbol{B}_k|}}{\sqrt{n_k}}\right)^{-1}\boldsymbol{H}_k\boldsymbol{L}_{p,k}, \tag{16}$$

$$\hat{\boldsymbol{x}}_k = \hat{\boldsymbol{x}}_{k|k-1} + (\boldsymbol{K}_k \otimes \boldsymbol{I}_d)\,\boldsymbol{g}_k, \tag{17}$$

$$\hat{v}_k = \hat{v}_{k|k-1} + n_k, \tag{18}$$

$$\boldsymbol{N}_k = L_{s,k}^{-\mathrm{T}} L_{s,k}^{-1}\boldsymbol{g}_k\boldsymbol{g}_k^{\mathrm{T}}, \tag{19}$$

$$\hat{\boldsymbol{V}}_k = \hat{\boldsymbol{V}}_{k|k-1} + \boldsymbol{N}_k + \boldsymbol{B}_k^{-1}\bar{\boldsymbol{Z}}_k\boldsymbol{B}_k^{-\mathrm{T}}, \tag{20}$$

$$\hat{\boldsymbol{X}}_k = \hat{\boldsymbol{V}}_k / \left(\hat{v}_k - 2d - 2\right). \tag{21}$$

Notice that we simplify the update of $\boldsymbol{P}_{k|k}$ by updating $\boldsymbol{L}_{e,k}$ in place of directly updating $\boldsymbol{P}_{k|k}$. Notably,

$$\boldsymbol{L}_{e,k}\boldsymbol{L}_{e,k}^{\mathrm{T}} = \boldsymbol{P}_{k|k-1} - \boldsymbol{K}_k S_k \boldsymbol{K}_k^{\mathrm{T}}, \tag{22}$$

which is consistent with the update of the 1-d kinematic error covariance matrix in [9].

The abstract level of the aforementioned DNN-aided HD-EOT enhanced predication and update procedure is summarized in Fig.1.

B. EOTNet Architecture

We now elaborate the network framework of the EOTNet, composed of perceptron models [21] and GRUs [22]. Specifically, the former handle basic matrix operations such as inversion and multiplication, whereas the latter estimate temporally dependent statistics [16].

To enforce constraints C1 and C2, we first design MLP-SPD modules generating lower triangular matrices with positive principal diagonal elements. Each module consists of an input, hidden, and output layer followed by a softplus activation [23]. The activation function maps each principal diagonal element from the output layer to a positive scalar.

We adopt a split structure in the EOTNet, dividing the GRU and multilayer perceptron (MLP) modules into two units corresponding to the prediction and update steps of the HD-EOT, as illustrated in Fig. 2.

The *Process Noise Unit (PNU)* estimates the square-root $\boldsymbol{L}_{p,k}$ of the predicted 1-d state error covariance $\boldsymbol{P}_{k|k-1}$ via a structured procedure. At each time step k, the 1-d process noise covariance $\boldsymbol{D}_k$ is obtained via a GRU module $\mathrm{GRU}_{\boldsymbol{D}}$ with the input feature $\Delta\hat{\boldsymbol{x}}_{k-1} \triangleq \hat{\boldsymbol{x}}_{k-1|k-1} - \hat{\boldsymbol{x}}_{k-1|k-2}$. Accordingly, $\boldsymbol{P}_{k|k-1}$ is predicted via the other GRU module $\mathrm{GRU}_{\boldsymbol{P}}$ with the input feature $\Delta\tilde{\boldsymbol{x}}_{k-1} \triangleq \hat{\boldsymbol{x}}_{k-1|k-1} - \hat{\boldsymbol{x}}_{k-2|k-2}$ and $\boldsymbol{D}_k$.

The *Measurement Noise Unit (MNU)* estimates both the square-root $L_{s,k}$ of the innovation covariance S_k and the observation distortion matrix $\boldsymbol{B}_k$. In the first stage, the intermediate statistics $\hat{\boldsymbol{X}}'_{k|k-1}$ and $\boldsymbol{R}_k$ are jointly estimated by a GRU module $\mathrm{GRU}_{\boldsymbol{XR}}$ with the input features $\tilde{\boldsymbol{Z}}_k \triangleq \frac{\bar{\boldsymbol{Z}}_k}{n_k - 1}$ and $\Delta\tilde{\boldsymbol{Z}}_{k-1} \triangleq \frac{\bar{\boldsymbol{Z}}_k}{n_k-1} - \frac{\bar{\boldsymbol{Z}}_{k-1}}{n_{k-1}-1}$. The obtained statistics, along with the input feature $\Delta\hat{\boldsymbol{Z}}_{k-1} \triangleq \frac{\bar{\boldsymbol{Z}}_{k-1}}{n_{k-1}-1} - \hat{\boldsymbol{X}}_{k-1|k-1}$, are fed into $\mathrm{GRU}_{\boldsymbol{B}}$ to estimate a relaxed distortion matrix $\boldsymbol{B}'_k$. Inposing the structural constraint C2, we have $\boldsymbol{B}_k$ via the refinement with an MLP-SPD module. In the second stage, S_k is estimated by GRU_S with the feature $\boldsymbol{g}_k$ and the scaled $\boldsymbol{B}_k$, and then $L_{s,k}$ is obtained via another MLP-SPD. Note that the above procedure closely resembles (4) and (8), respectively.

C. Training strategy

We train the proposed HD-EOT network in an end-to-end manner using a supervised learning framework. Specifically, given a labeled dataset $\mathcal{D} = \{(\mathbb{X}^{(b)}, \mathbb{Y}^{(b)})\}_{b=1}^{\mathcal{B}}$, where each trajectory b contains the ground-truth states of the EO and associated pre-processed measurements, i.e.,

$$\mathbb{X}^{(b)} \triangleq \left\{\mathcal{X}_0^{(b)}, \ldots, \mathcal{X}_k^{(b)}, \ldots, \mathcal{X}_{T_b}^{(b)}\right\}, \tag{23}$$

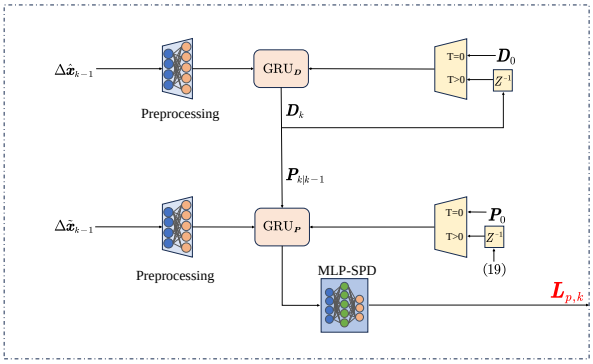
$$\mathbb{Y}^{(b)} \triangleq \left\{\mathcal{Y}_1^{(b)}, \ldots, \mathcal{Y}_k^{(b)}, \ldots, \mathcal{Y}_{T_b}^{(b)}\right\}, \tag{24}$$

where

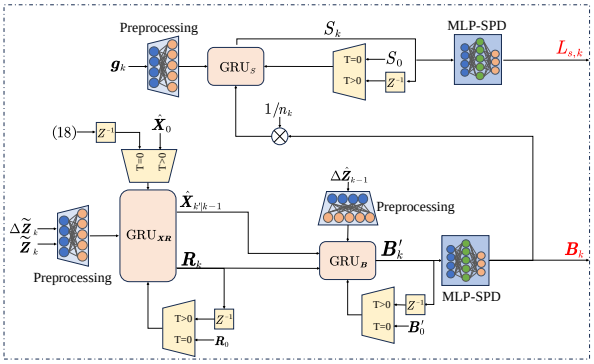
$$\mathcal{X}_k^{(b)} \triangleq \left\{\boldsymbol{x}_k^{(b)}, \boldsymbol{X}_k^{(b)}\right\}, \mathcal{Y}_k^{(b)} \triangleq \left\{\bar{\boldsymbol{z}}_k^{(b)}, \bar{\boldsymbol{Z}}_k^{(b)}\right\}, \tag{25}$$

respectively denote the pairs of the EO states and the measurements.

We propose to apply the square of the Gaussian Wasserstein distance (GWD) [24] as the loss function, which comprehensively captures both the instantaneous discrepancies between



(a) The PNU



(b) The MNU

Fig. 2: The architecture of the EOTNet.

the kinematic state estimate $\hat{\boldsymbol{x}}_k(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2)$ and its ground truth $\boldsymbol{x}_k$ and the extension state estimate $\hat{\boldsymbol{X}}_k(\boldsymbol{\Theta}_2)$ and its ground truth $\boldsymbol{X}_k$, i.e.,

$$\ell_{\mathrm{GWD}}(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2) \triangleq \|\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2)\|^2 + \mathrm{tr}\left\{\boldsymbol{X}_k + \hat{\boldsymbol{X}}_k(\boldsymbol{\Theta}_2) - 2\sqrt{\sqrt{\boldsymbol{X}_k}\hat{\boldsymbol{X}}_k(\boldsymbol{\Theta}_2)\sqrt{\boldsymbol{X}_k}}\right\}. \quad (26)$$

Notice that, the proposed GWD loss could be readily shown to be differentiable, enabling end-to-end gradient-based optimization.

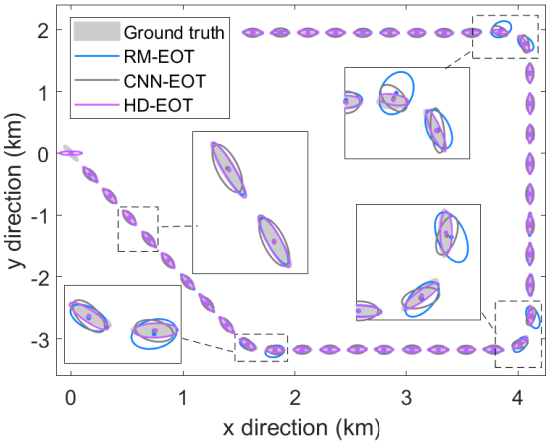


Fig. 3: The trajectories of the EO in a typical trial.

IV. SIMULATIONS AND RESULTS

We consider tracking a vessel, which is modeled as an ellipse with a major axis of 340 m and a minor axis of 80 m [3], [8], [9]. The initial kinematic state $\boldsymbol{x}_0$ of the EO is $[0\,\mathrm{m}, 0\,\mathrm{m}, 13.9\,\mathrm{m/s}, 13.9\,\mathrm{m/s}]$, corresponding to an approximate constant velocity (CV) of 27 knots. The EO then continues with CV motion and undergoes a series of maneuvers, including a 45° turn and two 90° turns at a constant turn rate (CT), with a fixed observation interval of $T = 5\,\mathrm{s}$.

At each time step k, n_k measurements are obtained from CFAR-processed radar data, where n_k follows a Poisson

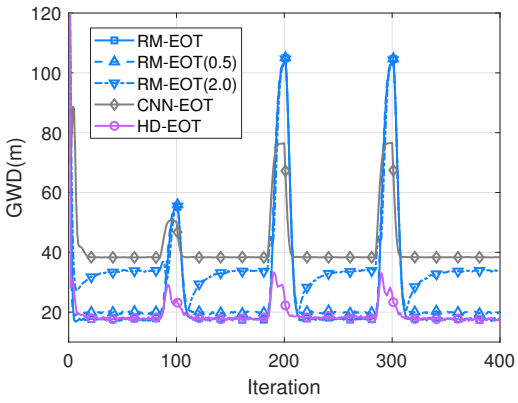


Fig. 4: The average GWDs of different approaches.

distribution with a mean of 50 [3], [8], [9]. We consider that the measurement noise covariance matrix and the process noise covariance matrix are respectively given by

$$\boldsymbol{R} = \begin{bmatrix} r^2 & 0 \\ 0 & r^2 \end{bmatrix} \mathrm{m}^2, \quad \boldsymbol{D} = \begin{bmatrix} q_1^2 & 0 \\ 0 & q_2^2 \end{bmatrix} \mathrm{m}^2, \quad (27)$$

where $r = 20$, $q_1 = 0.01$ and $q_2 = 0.001$. Additionally, the state transition matrix and the extension evolution matrix are given by $\boldsymbol{F}_k = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}$, $\boldsymbol{A}_k = \frac{1}{\sqrt{\delta_k}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, with the degree of freedom $\delta_k = 3$ of the Wishart distribution, for both the HD-EOT and RM-EOT.

For training the HD-EOT and the CNN-EOT, we construct the simulation dataset based on the above signal model by randomly generating the initial kinematic state and respectively scaling the lengths of the major and minor axes with random factors uniformly generated between 0.8 and 1.2.

At each time step k, we evaluate the comprehensive tracking performance averaged with $\mathcal{J} = 1000$ independent trials using

the GWD loss [24], the square of which is given by

$$L^2_{\mathrm{GWD}}(k) = \frac{1}{\mathcal{J}} \sum_{j=1}^{\mathcal{J}} \Bigg(\left\| \boldsymbol{x}_k^{(j)} - \hat{\boldsymbol{x}}_k^{(j)} \right\|^2 + \mathrm{tr} \left\{ \boldsymbol{X}_k^{(j)} + \hat{\boldsymbol{X}}_k^{(j)} - 2\sqrt{\sqrt{\boldsymbol{X}_k^{(j)}} \hat{\boldsymbol{X}}_k^{(j)} \sqrt{\boldsymbol{X}_k^{(j)}}} \right\} \Bigg), \quad (28)$$

where $\hat{\boldsymbol{x}}_k^{(j)}$ and $\hat{\boldsymbol{X}}_k^{(j)}$ respectively denote the estimates of the kinematic state $\boldsymbol{x}_k^{(j)}$ and extension state $\boldsymbol{X}_k^{(j)}$ in the jth trial.

We now consider evaluating the comprehensive tracking performance of the proposed DNN-aided HD-EOT methods, along with the RM-EOT [9] and the CNN-EOT [13]. The estimated trajectories of the EO in a typical trial and the average comprehensive tracking performance, are presented in Figs. 3 and 4, respectively. We further consider the RM-EOT (0.5) and RM-EOT (2.0), whose presumed average powers of the measurement noise are $0.5\times$ and $2.0\times$ ground-truth counterpart, respectively. The above RM-EOTs are observed to be significantly degraded, as the RM-EOT relies on accurate prior noise covariance, which is generally unavailable in practical X-band marine radar scenarios [1].

In contrast, the HD-EOT does not depend on any prior noise covariance knowledge. The proposed HD-EOT is observed to outperform all the competitors, especially during the turning maneuvers with partially unknown parameters. In tracking the EO of the CV motion, the performance of the HD-EOT is observed to be quite similar to that of the RM-EOT, as the HD-EOT strictly resembles the inference logic of the RM-EOT. In contrast, the computationally expensive CNN-EOT is observed to fail to accurately track the EO owing to the limited computational resources for training in the scenarios. On the other hand, the HD-EOT is observed to exhibit superior robustness compared to its competitors in tracking the EO of the CT motion, owing to the hybrid inference, which would potentially ameliorate the problem of the partially unknown parameters.

Further, we consider herein several simplified scenarios with different measurement noise covariances and parameter biases, with $T_b = 100$ data samples with the EO moving with a constant velocity. Due to inherent limitations of the CNN-EOT, we compare herein only the HD-EOT and RM-EOT. The steady-state comprehensive tracking performance over $\mathcal{J}$ trials are summarized in Table I.

TABLE I: The steady-state GWDs

r	Method	No Bias	a=10°	a=40°	f=10°	f=40°
10	RM-EOT	12.7275	13.1206	14.9826	13.5823	24.6223
	HD-EOT	**12.6353**	**12.6368**	**12.6409**	**12.6453**	**12.6404**
20	RM-EOT	13.0334	14.2652	16.1970	13.9512	24.8250
	HD-EOT	**12.9441**	**12.9440**	**12.9350**	**12.9349**	**12.9438**
30	RM-EOT	13.3693	15.4601	17.2063	14.3634	25.0763
	HD-EOT	**13.2745**	**13.2745**	**13.2746**	**13.2745**	**13.2746**
40	RM-EOT	13.6284	16.3725	17.8988	14.7175	25.3400
	HD-EOT	**13.6098**	**13.7098**	**13.7096**	**13.7098**	**13.7097**

The parameters a and f represent the biases in the tracking methods, respectively. Specifically, the biased state transition matrices are defined as

$$\boldsymbol{A}_{\mathrm{bias}} = \boldsymbol{\Omega}(a)\boldsymbol{A}\boldsymbol{\Omega}(a)^{\mathrm{T}}, \quad \boldsymbol{F}_{\mathrm{bias}} = \boldsymbol{\Omega}(f)\boldsymbol{F}. \quad (29)$$

where $\boldsymbol{A}$ and $\boldsymbol{F}$ denote the ground-truth state transition matrices of the EO motion. The proposed HD-EOT and RM-EOT incorporate these partially unknown parameters as the underlying state-space model parameters for inference. The former is observed to consistently outperform the latter across all tested measurement noise levels and parameter biases. When these partially unknown parameters are fed into both methods, the proposed HD-EOT consistently outperforms the RM-EOT, regardless of the noise environment or the severity of parameter bias. This further validates HD-EOT's strong robustness without requiring prior statistical noise knowledge.

TABLE II: Relative computation time and GPU memory usage

Method	Time Ratio	GPU Memory Ratio
CNN-EOT	18.51	75.70
HD-EOT	**0.43**	**1.02**

We further compare the relative computation time and GPU memory usage with respect to that of the RM-EOT over $\mathcal{J}$ trials. It could be observed in Table II that the proposed HD-EOT is computationally more efficient than its competitors, achieving over a $2\times$ speedup compared to the RM-EOT while maintaining almost the same memory consumption. On the other hand, the CNN-EOT demands over $18\times$ computation time and $75\times$ memory overhead due to its deep convolutional architecture. This would potentially result from the lightweight design of the HD-EOT where costly operations such as matrix inversion in the RM-EOT are replaced by MLPs. Nevertheless, the incorporation of the small-scale DNNs with the inference logic resembling the RM-EOT could markedly enhance the comprehensive tracking behavior of the HD-EOT at barely extra cost of the interpretability of the inference logic.

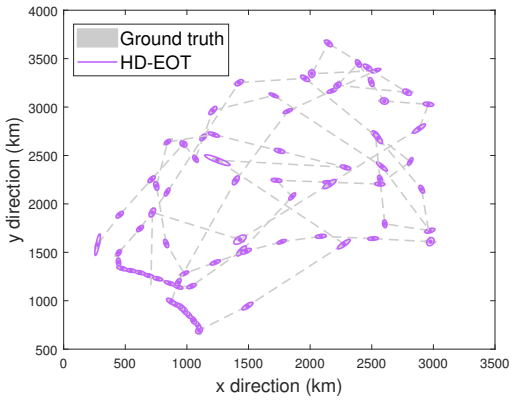


Fig. 5: The trajectories of the EO in the real-world dataset.

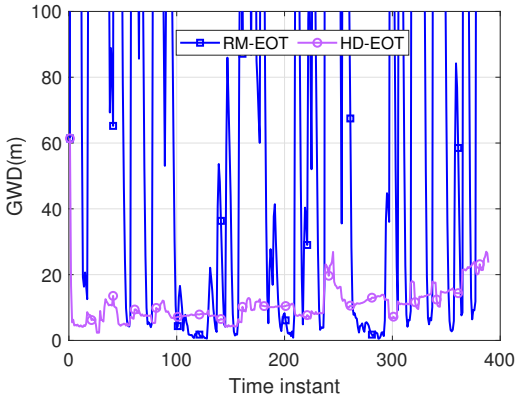


Fig. 6: The average GWDs of different approaches in the real-world dataset.

Finally, we assess the tracking performance of the proposed HD-EOT on a real-world X-band radar dataset reported by Liu et al. [25]. We extract the AIS trajectory of a small fast boat that was approximately ten meters long and perform linear interpolation to obtain a uniform $T = 30$s observation interval [25]. We then synthesize the radar measurements from these interpolated points. The measurement noise covariance is set to $R = \mathrm{diag}\{3,\, 3\}\,\mathrm{m}^2$. We adopt this processing workflow because the original radar measurements and the AIS ground truth suffer from significant spatio-temporal misalignment [25] and because CFAR-processed Cartesian observations often exhibit multi-target occlusions that lie beyond the scope of this study. The resulting true trajectory is shown in Fig. 5 over five observation intervals. Notice that we zoom in both the major and minor axes of the trajectory EO by a factor of twenty to enhance clarity.

To better capture realistic maneuvering behaviors and varying process noise, we enrich the diversity of our training dataset by including CT and constant-acceleration (CA) trajectories along with the original CV trajectories. We also introduce variability in the process noise covariance by randomly sampling the new intensity D_{new} from the values D, $5D$, $10D$ and $100D$, while keeping all the other state-space model and filter parameters unchanged.

From Figs. 5 and 6, it could be observed that the proposed HD-EOT method closely aligns with the interpolated AIS ground truth even during rapid maneuvers, whereas the RM-EOT method fails to track the EO, thereby demonstrating the robustness and efficacy of the HD-EOT and its superiority over the conventional RM-EOT algorithm on the challenging real X-band radar data.

V. CONCLUSION

In this work, we propose a lightweight and interpretable DNN-aided extended object tracking approach, the hybrid-driven EOT, which integrates several GRUs and MLPs into the random matrix EOT framework. To ensure that the obtained statistics from the GRUs remain positive definite and symmetric, we further develop a positive definiteness and symmetry preservation scheme based on the square-root filtering method and MLP-SPD modules. Both the simulation and experimental results generated from the pre-processed synthetic marine X-band radar measurements, validate the effectiveness of the proposed HD-EOT approach.

REFERENCES

[1] W. Huang, X. Liu, and E. W. Gill, "Ocean wind and wave measurements using X-band marine radar: A comprehensive review," *Remote Sens.*, vol. 9, no. 12, p. 1261, 2017.

[2] K. Granstrom, M. Baum, and S. Reuter, "Extended object tracking: Introduction, overview and applications," *arXiv preprint arXiv:1604.00970*, 2016.

[3] J. Koch, "Bayesian approach to extended object and cluster tracking using random matrices," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 3, pp. 1042–1059, Jul. 2008.

[4] S. Haag, B. Duraisamy, W. Koch *et al.*, "Radar and LIDAR target signatures of various object types and evaluation of extended object tracking methods for autonomous driving applications," in *2018 21st International Conference on Information Fusion (FUSION)*. IEEE, 2018, pp. 1746–1755.

[5] L. Mihaylova and W. Aftab, "Recent advances in extended and group objects tracking," in *Proc. Artif. Intell. Mil. Multisensor Fusion Engines (STO-MP-SET-262)*, 2018, pp. 14–1.

[6] L. Sun, J. Dong, and D. Gao, "Tracking of maneuvering extended object with random hypersurface model based on adaptive uncorrelated conversion filter," *Int. J. Aeronaut. Space Sci.*, pp. 1–15, 2024.

[7] L. Wang and R. Zhan, "Joint detection, tracking, and classification of multiple maneuvering star-convex extended targets," *IEEE Sens. J.*, vol. 24, no. 4, pp. 5004–5024, 2024.

[8] M. Feldmann, D. Franken, and W. Koch, "Tracking of extended objects and group targets using random matrices," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1409–1420, Apr. 2011.

[9] J. Lan and X. R. Li, "Tracking of extended object or target group using random matrix: New model and approach," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 6, pp. 2973–2989, Dec. 2016.

[10] L. Zhang and J. Lan, "Extended object tracking using random matrix with skewness," *IEEE Trans. Signal Process.*, vol. 68, pp. 5107–5121, 2020.

[11] ——, "Tracking of extended object using random matrix with non-uniformly distributed measurements," *IEEE Trans. Signal Process.*, vol. 69, pp. 3812–3825, 2021.

[12] L. Jiao, D. Wang, Y. Bai *et al.*, "Deep learning in visual tracking: A review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 9, pp. 5497–5516, 2021.

[13] S. Steuernagel, K. Thormann, and M. Baum, "CNN-based shape estimation for extended object tracking using point cloud measurements," in *Proc. Int. Conf. Inf. Fusion (FUSION)*, 2022, pp. 1–8.

[14] B. Bian and H. Chen, "Convolutional neural network for ellipse extended target tracking," *Chin. J. Inf. Fusion*, vol. 1, no. 2, pp. 93–108, 2024.

[15] N. Shlezinger, J. Whang, Y. C. Eldar, and A. G. Dimakis, "Model-based deep learning," *Proc. IEEE*, vol. 111, no. 5, pp. 465–499, 2023.

[16] G. Revach, N. Shlezinger, X. Ni *et al.*, "Kalmannet: Neural network aided Kalman filtering for partially known dynamics," *IEEE Trans. Signal Process.*, vol. 70, pp. 1532–1547, 2022.

[17] G. Choi, J. Park, N. Shlezinger *et al.*, "Split-KalmanNet: A robust model-based deep learning approach for state estimation," *IEEE Trans. Veh. Technol.*, vol. 72, no. 9, pp. 12 326–12 331, 2023.

[18] K. Granström, A. Natale, P. Braca *et al.*, "Gamma Gaussian inverse Wishart probability hypothesis density for extended target tracking using X-band marine radar data," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6617–6631, 2015.

[19] J. Lan, "Extended object tracking using random matrix with extension-dependent measurement numbers," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 4, pp. 4464–4477, 2023.

[20] A. Andrews, "A square root formulation of the Kalman covariance equations," *AIAA J.*, vol. 6, no. 6, pp. 1165–1166, 1968.

[21] K.-L. Du, C.-S. Leung, W. H. Mow *et al.*, "Perceptron: Learning, generalization, model selection, fault tolerance, and role in the deep learning era," *Mathematics*, vol. 10, no. 24, p. 4730, 2022.
[22] J. Chung, C. Gulcehre, K. Cho *et al.*, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.
[23] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. Int. Conf. Artif. Intell. Stat. (AISTATS)*, 2011, pp. 315–323.
[24] S. Yang and M. Baum, "Tracking the orientation and axes lengths of an elliptical extended object," *IEEE Transactions on Signal Processing*, vol. 67, no. 18, pp. 4720–4729, 2019.
[25] N. Liu, H. Ding, Y. Huang *et al.*, "Annual progress of sea-detecting x-band radar and data acquisition program," *J. Radars*, vol. 10, no. 1, 2021.