IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 25 October 2023, accepted 27 November 2023, date of publication 7 December 2023,
date of current version 14 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3340986

RESEARCH ARTICLE

ConvLSTM-Based Vehicle Detection and Localization in Seismic Sensor Networks

ERDEM KÖSE AND ALİ KÖKSAL HOCAOĞLU
Department of Electronics Engineering, Gebze Technical University, Gebze, 41400 Kocaeli, Turkey
Corresponding author: Ali Köksal Hocaoğlu (khocaoglu@gtu.edu.tr)

ABSTRACT Localization of moving military vehicles plays a vital role for border security and safeguarding high-security facilities. Commonly applied range-based localization techniques such as time of arrival, time difference of arrival, angle of arrival, and received signal strength rely on known transmitters. However, when seismic sensor networks are used for localization of moving targets, where moving targets can be treated as unknown transmitters. In this work, we consider a scenario where only receivers are deployed to perceive seismic signals transmitted by the moving military vehicles with unknown locations. Consequently, conventional closed-form equations for distance-based trilateration are not applicable. To address this challenge, we present a novel approach for accurate localization. Our method involves clustering closely deployed sensor nodes to effectively fuse their information to estimate the positions of the moving military vehicles. We leverage multiple-input convolutional neural networks, utilizing one input to represent the short-time discrete Fourier transform of signals from each node, and another input to encode the relative locations of sensors within clusters. Through extensive experimentation, we demonstrate that our proposed method significantly reduces localization errors when compared to existing distributed regression methods.

INDEX TERMS Convolutional neural networks, long short-term memory, military vehicles, seismic waves, vehicle detection, vehicle location estimation, wireless sensor networks.

I. INTRODUCTION

Detection, classification, and localization of moving ground vehicles in battlefield or high-security facilities play a crucial role in remote sensing tasks [1], [2]. While multimedia surveillance devices, such as various types of cameras, have been commonly used for these purposes [3], [4], there are now more cost-effective alternatives available, such as acoustic, seismic, and magnetic sensors [5]. These sensors offer significant improvements in signal-to-noise ratio, accuracy, and cost, making detection, classification, and localization in wireless sensor networks utilizing these technologies have been increasingly popular in recent times.

Seismic sensor networks are common in usage for early earthquake warning [6], railway condition monitoring [7], vehicle speed estimation [8]. Tracked and trackless moving ground vehicle discrimination in seismic sensor networks [9] is one of the most popular and improving studies. The third Sensor Information Technology (SensIT) Situational Experiment of the Defense Advanced Research Projects Agency (DARPA) [9] generated one of the most important datasets called SensIT situational EXperiment (SITEX02) about military vehicles in seismic and acoustic distributed sensor networks. SITEX02 dataset contains seismic and acoustic signals from two types of armored vehicles, which are fully tracked assault amphibious vehicle (AAV) and fully wheeled dragon wagon (DW). The Defense Advanced Research Projects Agency (DARPA) program SensIT is based on the concept of detecting and identifying targets at sensor field to support remote situation awareness capabilities. This dataset is very popular in the literature for classification and detection; but it also contains the ground truth trajectory data recorded via GPS (Global Positioning System) transmitters. Therefore, it is also a very good dataset to benchmark localization algorithms.

Vehicle detection is usually the first step of location estimation and one of the most popular methods is Constant False Alarm Rate (CFAR) algorithm [10]. Duarte and Hu

The associate editor coordinating the review of this manuscript and approving it for publication was Guangjie Han.

used k-Nearest Neighbors (kNN) [9] for event detection in military environment. Kalra et al. investigated performance of Smooth Pseudo Wigner-Ville Distribution (SPWVD) [11] on detection of moving military vehicles. Bin et al. compared Seismic Fractal Features with short/long time average ratio (STA/LTA) and fractal dimension-based support vector machine (FD-SVM) [12] for ground moving target detection.

After detection step, there are different approaches for vehicle localization [13]. Most popular range-based ones are Angle of Arrival (AOA), Direction of Arrival (DOA), Time Difference of Arrival (TDOA) and Received Signal Strength Index (RSSI). Boettcher et al. proposed acoustic TDOA algorithm [14] for localization with SITEX00 dataset. Li et al. suggested an energy-based localization algorithm for the same dataset [15]. Rahman et al. used Optimized Maximum Likelihood for localization via acoustic modality in SITEX02 dataset [16]. Le Borgne et al. investigated centroid scheme and distributed regression for localization in SITEX02 dataset [17]. Apart from SITEX02 dataset, Gouda et al. [18] used Inertial Measurement Unit (IMU) with Convolutional Neural Networks (CNN) to estimate robot locations in evenly distributed sensor networks. Also, Yuan et al. [19] proposed Passive Radio Frequency (PRF) distribution for 3-D indoor target localization.

It is well known that a sensor cluster with at least three sensors is required to estimate target location accurately in sensor network environment, also there are studies that show clustering the nodes improves classification performance too. Taheri et al. used time-varying autoregressive model (TVAR) with clustering nodes to improve target identification performance [20]. Zwartjes et al. is proposed QUantile Estimation after Supervised Training (QUEST) [21] for adaptive learning to classify military vehicles.

In this paper, we propose a novel approach to detect and estimate target positions using closely deployed seismic sensor nodes with reduced localization error. Instead of conventional trilateration techniques, we leverage the effectiveness of Convolutional Long-Short Term Memory (ConvLSTM) [22] due to the limitations of seismic signal energy compared to RSSI, as shown by Le Borgne et al. [17]. The use of seismic data from the SITEX02 dataset allows us to train and validate our location estimation model. By clustering and fusing closely deployed seismic sensor nodes and utilizing narrow bandwidth features from the frequency domain, we achieve low localization error, deviating from the conventional approach of using all raw acoustic and seismic data in the field, as demonstrated by Le Borgne et al. [17]. While terrain analysis could be valuable, the SITEX02 dataset, lacks terrain data. Therefore, the influence of terrain on the reliability and effectiveness of seismic sensors is out of scope for the study.

The rest of paper is organized as follows: in Section II, the proposed method is given. Section III introduces the network model. Performance criteria, test results and comparison with previous works are given in Section IV. Lastly, the fifth section gives a summary and concludes the work.

II. PROPOSED METHOD

Trilateration serves as a fundamental approach for range-based localization and forms the basis of our method. Circular trilateration, illustrated in Fig. 1, involves three sensors, namely S_1, S_2, and S_3, situated at distinct locations and r_1, r_2 and r_3 are the sensor-target distances. Various methods such as AOA, TOA, TDOA, or RSSI are employed to estimate the location of the target *tar* by processing the signals from these sensor nodes [23].

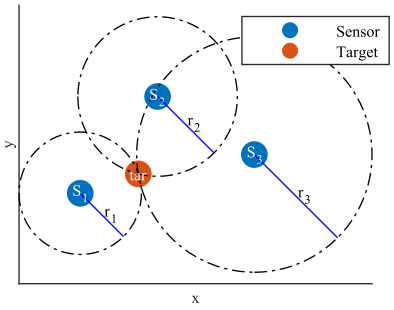

FIGURE 1. Circular trilateration.

The relation between actual target position $\mathbf{x}_{\text{tar}} = [x_{\text{tar}}, y_{\text{tar}}]^{\text{T}}$ and the sensor position $\mathbf{x}_{s_i} = \left[x_{s_i}, y_{s_i}\right]^{\text{T}}$ is given in (1)

$$\left(x_{s_i} - x_{\text{tar}}\right)^2 + \left(y_{s_i} - y_{\text{tar}}\right)^2 = r_i^2 \tag{1}$$

Sensor locations x_{s_i} and y_{s_i} are known variables. If the sensor-target distance r_i is also known, the target's position can be determined using a linear equation system constructed with (1) for $i = 1, 2, 3$ [13].

If r_i is not known, another information that has linear relationship with r_i must be known; thus, the triliteration is mostly used with RSSI of the mobile networks [24]. Using this idea, Le Borgne et al. [17] used the energy of seismic signals instead of RSSI and showed that trilateration with seismic energy itself is not useful. Because there is no closed form RSSI and distance functions for seismic signals and signal pattern changes by velocity, acceleration, soil and vehicle type. Therefore, they followed a different approach and used all the sensor nodes in the SITEX02 by applying distributed regression to raw data to achieve better results.

We replaced traditional methods with a ConvLSTM layer [22] to leverage vehicle position changes over time and their impact on seismic frequency domain features. This approach integrates Convolutional and LSTM layers into a unified framework, extending the feature space into spatio-temporal dimensions. This integration improves upon separate Convolutional and LSTM models, enhancing the capabilities of data-driven methods. The model takes frequency domain features of raw seismic signal and seismic sensor locations as separate inputs, producing target locations relative to the center of clustered sensor nodes as the output.

The dataset SITEX02 has a good node deployment scheme for target localization. Therefore, we will use it to explain how to cluster and extract features to localize targets using only three sensor nodes deployed in the field. We will first explain the dataset and then discuss the details of the proposed localization algorithm.

A. DATASET

The SITEX02 dataset [9] contains three different scenarios for vehicle route. GPS recordings for these events are available for validating the location estimations. Fig. 2 shows node deployment scheme for SITEX02 dataset. There are 23 nodes having location information. There are 9 trials for AAV, and there are 11 trials for DW. Nodes 6, 60 and 61 have defective data and missing detection information [21]. Due to budget constraints, we relied solely on the SITEX02 dataset for our study, preventing us from exploring practical deployment challenges in seismic sensor networks.

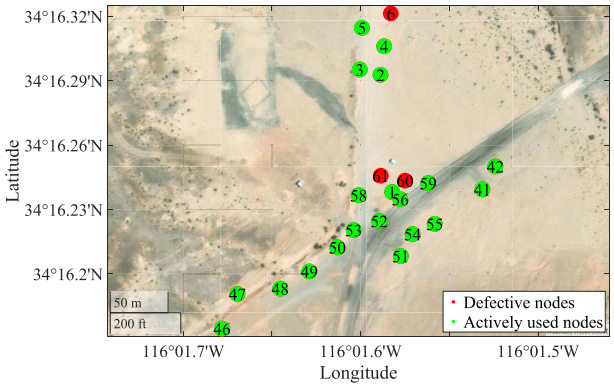

FIGURE 2. Node deployment scheme for SITEX02 dataset.

Each node has acoustic, seismic and passive infrared modalities. Data had been collected with sampling rate $f_s = 4960$ Hz for time blocks of $T_b = 0.75$ s. Each time block has $N = T_b f_s = 3720$ samples from each sensor modality.

B. FEATURES

1) FREQUENCY DOMAIN

Let $u[n]$ be the seismic signal with a duration of T_b seconds of an event. We calculate its N-point discrete Fourier transform (DFT) $U[k]$ using equation (2).

$$U[k] = \sum_{n=0}^{N-1} u[n] e^{-j\frac{2\pi k}{N} n} \tag{2}$$

Tensor of frequency domain features has three dimensions: First dimension is time index m, second one is sensor number i and third one is frequency index k, where $0 \leq K_1 \leq k \leq K_2 < N$. m is sequential time block index for giving temporal information about occurring event to CNN, where $0 \leq m < M$. We can represent them in three-dimensional tensor $U_{m,i}[k]$.

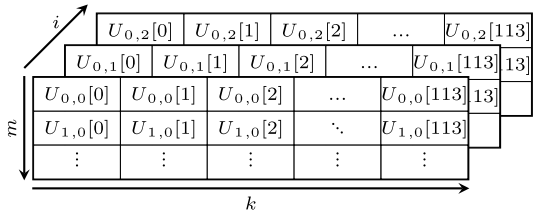

FIGURE 3. Tensor for frequency domain input.

Fig. 3 shows an example tensor structure for three sensor nodes, two time blocks, and frequency domain features with indices between 0 and 113. The reason for limiting the frequency index is that most of the energy of the seismic signal is between 0 and 150 Hz [25].

2) SENSOR LOCATION

We cluster sensor nodes suitable for triangulation. During this phase, it is also important to form clusters at far regions to represent different parts of the sensor deployment scheme. We then calculate the normalized locations according to the mean location of all sensors. Triangulated nodes can be seen in Fig. 4. The four clusters contain the sensor nodes $S_1 = \{s_5, s_3, s_4\}$, $S_2 = \{s_{47}, s_{46}, s_{48}\}$, $S_3 = \{s_{58}, s_{53}, s_{52}\}$ and $S_4 = \{s_{59}, s_{55}, s_{41}\}$ and $\mathbf{S} = \{S_j | j = 0, 1, 2, 3\}$.

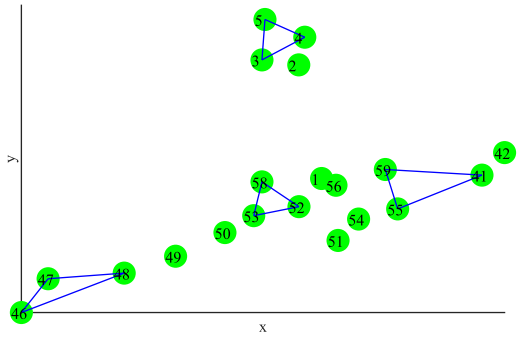

FIGURE 4. Selected triangulated node clusters based on target courses.

We use the relative locations of the sensors instead of using them directly to make it insensitive to geographic position of the site. We calculate the relative locations by subtracting the mean location of sensor cluster set as in equation (3).

$$\tilde{\mathbf{x}}_{s_i} = \begin{bmatrix} \tilde{x}_{s_i} \\ \tilde{y}_{s_i} \end{bmatrix} = \begin{bmatrix} x_{s_i} \\ y_{s_i} \end{bmatrix} - \mu_{\mathbf{x}}^{(\mathbf{S})} \tag{3}$$

where $\mu_{\mathbf{x}}^{(\mathbf{S})}$ is the center of the cluster set $\mathbf{S}$ and can be calculated using (4).

$$\mu_{\mathbf{x}}^{(\mathbf{S})} = \frac{1}{\sum_{S_j \in \mathbf{S}} |S_j|} \sum_{S_j \in \mathbf{S}} \sum_{s_i \in S_j} \begin{bmatrix} x_{s_i} \\ y_{s_i} \end{bmatrix} \tag{4}$$

Tensor of sensor location features has two dimensions: First dimension is sensor number i and second one axis index l, where $l = 0$ corresponds to the x-axis, and

$l = 1$ to the y-axis. We can represent them in two-dimensional tensor $V_{i,l}$.

Fig. 5 shows an example tensor structure for two coordinates and three sensor nodes constructed from $S_1 = \{s_5, s_3, s_4\}$ in a cluster.

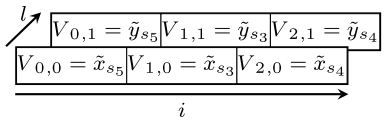

FIGURE 5. Tensor for sensor location input.

C. OUTPUT FORMAT

Training and inference of our algorithm depends on detection of target in sensor cluster S_j, which we call activated. And we use $\tilde{\mathbf{x}}_{s_i} \in S_j$ as sensor location input $V_{i,l}$ beside feature input $U_{m,i}[k]$ for both training and inference steps.

We make ground truth target location $\mathbf{x}_{\text{tar}} = [x_{\text{tar}}, y_{\text{tar}}]^{\text{T}}$ zero mean in training phase for generalization among sensor clusters. Target location $\tilde{\mathbf{x}}_{\text{tar}}$ relative to the center of activated sensor cluster S_j can be calculated using (5).

$$\tilde{\mathbf{x}}_{\text{tar}} = \begin{bmatrix} \tilde{x}_{\text{tar}} \\ \tilde{y}_{\text{tar}} \end{bmatrix} = \begin{bmatrix} x_{\text{tar}} \\ y_{\text{tar}} \end{bmatrix} - \mu_{\mathbf{x}}^{(S_j)} \tag{5}$$

where $\mu_{\mathbf{x}}^{(S_j)}$ is the center of the cluster S_j and can be calculated using (6).

$$\mu_{\mathbf{x}}^{(S_j)} = \frac{1}{|S_j|} \sum_{s_i \in S_j} \begin{bmatrix} x_{s_i} \\ y_{s_i} \end{bmatrix} \tag{6}$$

During inference or test phase, output of algorithm $\tilde{\mathbf{x}}_{\text{est}} = \left[\tilde{x}_{\text{est}}, \tilde{y}_{\text{est}}\right]^{\text{T}}$ can be denormalized using (7) to calculate estimated location in activated sensor cluster S_j.

$$\mathbf{x}_{\text{est}} = \begin{bmatrix} x_{\text{est}} \\ y_{\text{est}} \end{bmatrix} = \begin{bmatrix} \tilde{x}_{\text{est}} \\ \tilde{y}_{\text{est}} \end{bmatrix} + \mu_{\mathbf{x}}^{(S_j)} \tag{7}$$

III. NETWORK ARCHITECTURE

Our aim is to estimate the location of a target using frequency domain features and sensor locations. To achieve this, we leverage CNNs, enhancing their capabilities with ConvLSTM [22]. ConvLSTM combines convolutional operations with LSTM cells, enabling efficient processing of spatio-temporal data. Unlike traditional LSTM, which only considers temporal dependencies, ConvLSTM simultaneously captures spatial and temporal dependencies. This effectiveness makes it ideal for video analysis, action recognition, and other spatio-temporal data applications, where understanding both spatial and temporal dynamics is crucial for accurate predictions and feature learning. In this work, we use ConvLSTM to capture time-frequency features of seismic signatures of moving vehicles in a sensor network setting.

Fig. 6 illustrates our proposed mechanism, while Fig. 7 depicts the base model of our proposed method for detection and localization. *SeismicFourier* block uses frequency

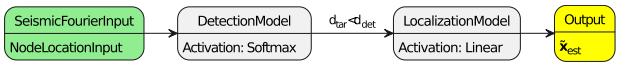

FIGURE 6. Detection-localization model flow.

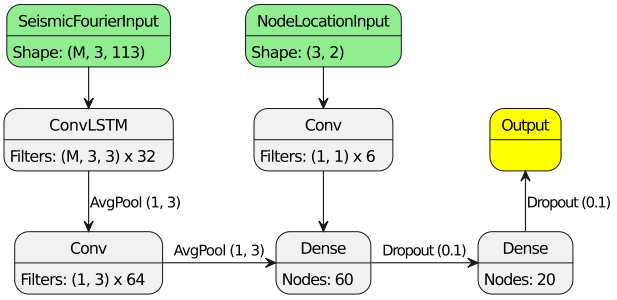

FIGURE 7. Overall network model for detection and localization.

domain features. We started with $M = 3$ temporal length ConvLSTM with $(3, 3)$ shape 32 filters and continued with $(1, 3)$ shape 64 filters similar to Darknet-19 [26]. We used $(1, 3)$ shape average pooling layers after each convolution layer. This configuration was selected because it is the smallest yet highest performance we achieved by our trials.

Activation functions except ConvLSTM are Leaky Rectified Linear Units (LReLU) [27] given in (8), where z is a variable from network. We set $\alpha = 0.001$.

$$\mathrm{f}_{\alpha}(z) = \begin{cases} \alpha z, & z < 0 \\ z, & z \geq 0, \end{cases} \quad \mathrm{f}'_{\alpha}(z) = \begin{cases} \alpha, & z < 0 \\ 1, & z \geq 0 \end{cases} \tag{8}$$

The base model defined so far is used for target detection and localization models, except the output layers. Detection model has cross-entropy in (9) as loss function, which is suitable for Softmax activation [28] in (10). We use loss function as $\mathrm{L}_{\text{CE}}\left(\mathrm{f}_{\text{SM}}(\mathbf{z}), \tilde{\mathbf{z}}\right)$ for the output layer.

$$\mathrm{L}_{\text{CE}}\left(\mathbf{z}, \tilde{\mathbf{z}}\right) = -\tilde{\mathbf{z}}^{\text{T}} \log(\mathbf{z}) \tag{9}$$

$$\mathrm{f}_{\text{SM}}(\mathbf{z}) = \frac{e^{\mathbf{z}}}{\sum_{z \in \mathbf{z}} e^{z}} \tag{10}$$

In localization model, the choice of the loss function is crucial. While classification-based options are limited and not applicable, various regression loss functions are available, such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Squared Logarithmic Error (MSLE), Cosine Similarity, Log-cosh, and Huber [29]. Among these, we select the Huber loss function due to its adaptability against outliers, setting it apart from the others. The Huber loss L_{δ} is defined in (11), where $\mathbf{z}$ is a variable from network, $\tilde{\mathbf{z}}$ is desired value and δ serves as a parameter to handle outliers. Since we use linear activation for output layer, our loss for the output layer is $\mathrm{L}_{\delta}\left(\mathbf{z}, \tilde{\mathbf{z}}\right)$.

$$\mathrm{L}_{\delta}\left(\mathbf{z}, \tilde{\mathbf{z}}\right) = \sum_{\substack{z \in \mathbf{z} \\ \tilde{z} \in \tilde{\mathbf{z}}}} \begin{cases} \frac{1}{2}(z - \tilde{z})^2, & |z - \tilde{z}| \leq \delta \\ \delta\left(|z - \tilde{z}| - \frac{1}{2}\delta\right), & |z - \tilde{z}| > \delta \end{cases} \tag{11}$$

IV. RESULTS

We used the seismic data of SITEX02 dataset, which consists of signals from AAV and DW vehicle types. AAV4, DW4, AAV5, DW5, AAV6 and DW6 trials are used for testing and rest for training. This means splitting the dataset which has 11700 samples to 70% training 30% testing portions.

We employed the Keras deep learning library [30]. The model training settings were configured as follows: We utilized the Adam optimizer [31] with a cyclical learning rate schedule [32], ranging from 1×10^{-5} to 1×10^{-2} with no decay and a step size of 100 for the detection model. For the localization model, we employed the Stochastic Gradient Descent optimizer [33] with a cyclical learning rate schedule ranging from 1×10^{-4} to 1×10^{-1} with a decay function of 0.9^{lr}, where lr represents the learning rate, applied after each cycle, along with a 50-step size. The detection model underwent 50 epochs, while the localization model underwent 200 epochs. The batch size for the detection model was set at 1000 training samples, and for the localization model, it was 20 training samples after the detection phase. We shuffled data, trained network ten times and gave the average results.

We investigated the effects of detection method and sensor configuration on performance in Table 1. Detection type *Original* is detected target information from [9] which is based on Constant False Alarm Rate (CFAR). Detection type *Distance Based* is the model we trained to estimate x_{tar}. As shown in Fig. 6, our model ignores all measurements that have $d_{\text{tar}} > d_{\text{det}}$ in (12). We inspired from the work of Duarte et al [34] which is used $d_{\text{det}} = 50$ m for distance-based decision fusion. We also investigated the effect of the number of sensor nodes in clusters. The results show that having more sensor nodes in clusters reduces localization error. Another conclusion we can draw from Table 1 is that during the training, ignoring the sensors far from the target has a significant impact on decreasing estimation error.

$$d_{\text{tar}} = \mathrm{d}_2\left(\mathbf{x}_{\text{tar}}, \mu_{\mathbf{x}}^{(S_j)}\right) \tag{12}$$

TABLE 1. Results in different sensor configurations.

Detection Method	Sensor Count in Cluster	Localization error	
		μ_{loc} (meters)	σ_{loc} (meters)
Original	1	41.1	30.2
Original	2	23.7	21.5
Original	3	15.3	14.0
Distance Based	1	23.2	16.4
Distance Based	2	14.6	10.7
Distance Based	3	9.7	5.9

We investigated whether to impose a maximum distance between target and sensor nodes d_{det} during training to see its impact on performance in Table 2. We chose F1 score, given in (13) as detection performance indicator because data is imbalanced. Equation (13) is built on true positive (TP), false positive (FP) and false negative (FN) counts. The results show that 50 m is one of the best values for d_{det} and verify the work of Duarte et.al. [34].

TABLE 2. Results in different detection distances.

d_{det} (meters)	F1_{det}	μ_{loc} (meters)	σ_{loc} (meters)
10	0.79	9.86	8.81
20	0.89	9.48	6.76
30	0.91	9.47	7.02
40	0.94	9.50	6.72
50	0.94	9.70	5.91
60	0.94	9.95	6.42
70	0.94	11.10	9.01
80	0.93	13.42	15.24
90	0.93	13.96	13.38
100	0.93	15.26	16.13

$$\text{F1}_{\text{det}} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \tag{13}$$

Temporal length, denoted as M, is an important parameter that requires consideration. In Fig. 8, we present results for different temporal lengths. Based on this analysis, we found that $M = 3$ yields the optimum performance, representing the sweet spot. The average inter-sensor distances are 29.95 m for S_1, 43.68 m for S_2, 27.36 m for S_3, and 44.52 m for S_4 (as shown in Fig. 4). Additionally, the average velocities of AAV and DW for each trial can be found in Table 3. Considering an average velocity of 7.09 m/T_b, selecting M values greater than 5 would result in the vehicle traveling more than 30 m, which exceeds the minimum inter-sensor distances. As a consequence, Fig. 8 illustrates a declining performance trend in such scenarios. If the speed of target increases, the parameter M should be reduced accordingly to maintain the localization performance.

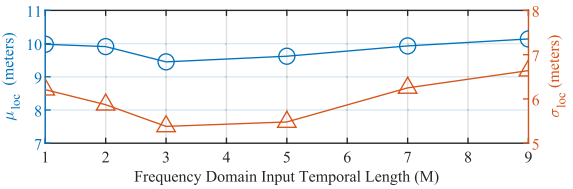

FIGURE 8. LSTM temporal length performance for $d_{\text{det}} = 50$ m.

TABLE 3. Vehicle velocity estimations from GPS data.

	AAV4	DW4	AAV5	DW5	AAV6	DW6
μ_{vel} (m/T_b)	6.38	6.98	7.46	7.61	7.08	7.04
μ_{vel} (km/h)	30.63	33.54	35.82	36.54	33.98	33.79
σ_{vel} (km/h)	3.28	1.75	1.15	0.60	0.69	0.81

The proposed method achieved the best performance with a localization error of 9.70 m and a standard deviation of 5.91 m. This represents a significant improvement compared to the centroid scheme and distributed regression methods proposed by Le Borgne et al. [17] (see Table 4). In addition, the Symbolic Dynamic Filtering algorithm proposed by Sindhu et al. [35] is intended for tracking, but it tracks a

TABLE 4. Methods in literature from localization aspect.

Method	μ_{loc} (meters)	σ_{loc} (meters)
Trilateration with Energy [17]	175.0	168.0
Centroid [17]	30.0	19.0
Regression [17]	22.0	14.0
Proposed Method	9.7	5.9

particular vehicle that has passed the given sensor node, not the position. Specifically, the average tracking rate is reported as 82.4%, making it challenging to make meaningful comparisons with our algorithm. Another research is using Classification and Localization using Estimated Dynamics and Multimodal data (CLEDM) by Lee et al. [36], but they used another dataset from US Army Research Laboratory [37], which is not publicly available.

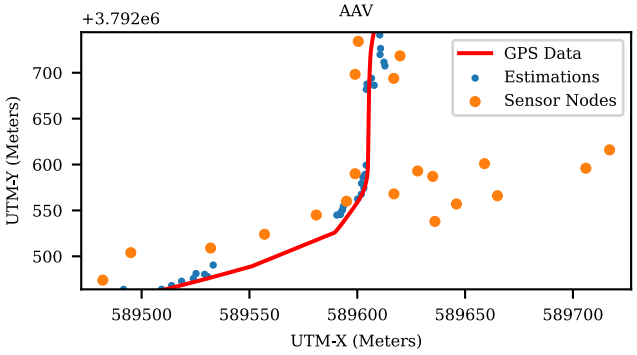

FIGURE 9. AAV4 target location estimation.

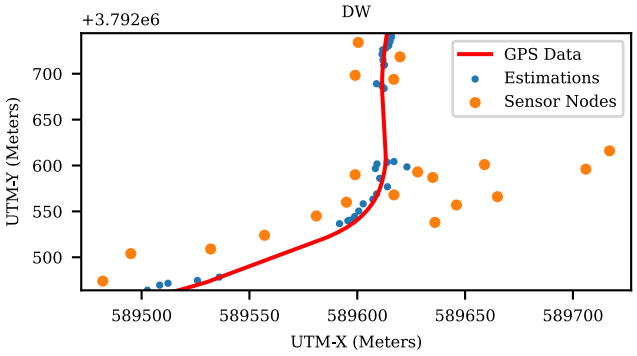

FIGURE 10. DW4 target location estimation.

We visualized the estimated paths for two different vehicles, AAV, and DW, along three distinct paths in Fig. 9, 10, 11, 12, 13 and 14. The figures indicate that AAV5, DW5 and DW6 exhibit bias in position estimations, but their trajectories are accurately represented. These observations are further supported by the statistical results presented in Table 5. In general, the model performs better for AAV, except for a faulty result for DW5. However, it is worth noting that the trajectory for DW5 aligns correctly with the path in Fig. 12.

TABLE 5. Trial statistics.

	AAV4	DW4	AAV5	DW5	AAV6	DW6
μ_{loc} (m)	7.12	7.50	10.44	17.19	8.65	10.64
$\text{median}_{\text{loc}}$ (m)	6.09	6.74	9.25	15.65	7.05	10.78
σ_{loc} (m)	4.26	3.89	6.30	5.37	7.10	3.13

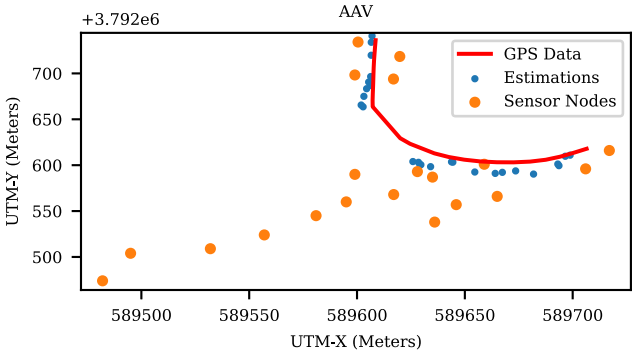

FIGURE 11. AAV5 target location estimation.

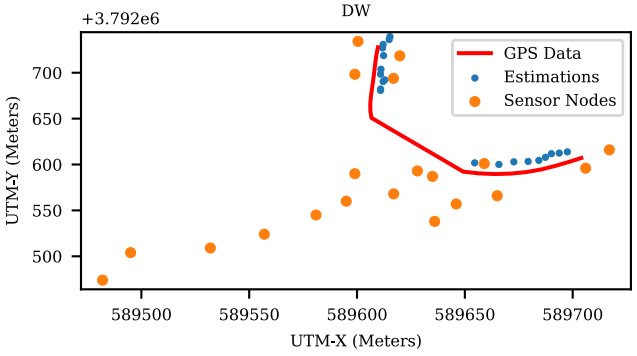

FIGURE 12. DW5 target location estimation.

The localization errors presented assume that GPS measurements of the dataset are accurate. However, for this dataset, the accuracy of GPS measurements is around 10 meters [38] and some sensor nodes have performance issues [21]. In addition to these: AAV vehicles have 3.28 m width and 8.15 m length [39], and DW vehicles have 3.32 m width and 7.72 m length [40]. AAV and DW dimensions show that vibration source has an average of 3.3 m width and 7.5 m length. Considering these drawbacks, the performance of the proposed algorithm is good for this dataset.

Another aspect of performance evaluation pertains to model inferencing. Both the detection and regression models employ a 32-bit floating-point data type, encompassing 48 mega Multiply-Accumulate operations (MMACs), with a size of 700 kB, and a memory consumption of 3.5 MB. These results in a runtime of 4.2 ms on a single-core AMD Ryzen 7 7745HX. Consequently, model inferencing requires 8.4 ms for each $T_bM = 2.25$ s block, making our model suitable for real-time applications. The average vehicle speed is 9.45 m/s, equivalent to 34 km/h, and the model operates every 2.25 seconds, allowing us to capture movements with a maximum location change of 19 m within the 50 m capture

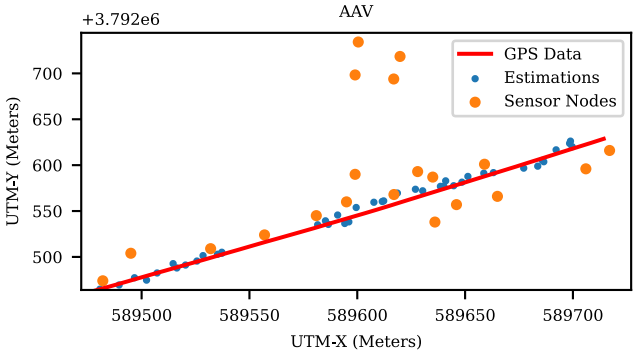

FIGURE 13. AAV6 target location estimation.

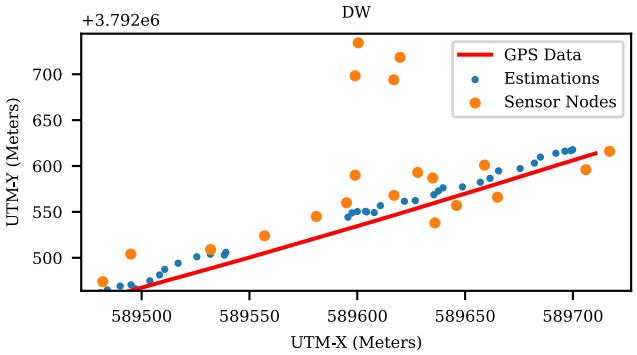

FIGURE 14. DW6 target location estimation.

range. Furthermore, feature extraction involves $N \log_2 N = 49152$ Multiply-Accumulate operations (MACs), taking a mere 52 μs on the AMD Ryzen 7 7745HX, which is negligible when compared to the model inferencing time. In contrast to the resource-intensive You Only Look Once version 3 (YOLOv3) [41], a well-known model deployable on mobile devices [42], our proposed approach is significantly less complex, requiring only about 33,000 MMACs. As a result, our approach is well-suited for running even on mobile devices.

V. CONCLUSION

We propose an alternative technique to trilateration and employ sensor locations along with spatio-temporal seismic frequency features as inputs to ConvLSTM for detecting and estimating military vehicle locations in a distributed seismic sensor network setting. Our investigations reveal that the number of sensor nodes in clusters significantly impacts performance, as does the detection distance. Compared to the regression method with the original detection scheme, our algorithm achieves an increase of 30% in localization accuracy. Moreover, using a 50 m detection range leads to more than 50% decrease in both μ_{loc} and σ_{loc}.

Furthermore, our findings demonstrate that utilizing only the spatio-temporal features from seismic sensors clustered into three independent groups reduces the localization error by more than 50%, in contrast to the approach involving all the acoustic and seismic sensors with raw temporal data in the network.

REFERENCES

[1] G. Jin, B. Ye, Y. Wu, and F. Qu, "Vehicle classification based on seismic signatures using convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 628–632, Apr. 2019.
[2] K. Bin, J. Lin, and X. Tong, "Edge intelligence-based moving target classification using compressed seismic measurements and convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
[3] K. Liu and G. Mattyus, "Fast multiclass vehicle detection on aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 9, pp. 1938–1942, Sep. 2015.
[4] S. Sivaraman and M. M. Trivedi, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1773–1795, Dec. 2013.
[5] M. Bernas, B. Placzek, W. Korski, P. Loska, J. Smyla, and P. Szymala, "A survey and comparison of low-cost sensing technologies for road traffic monitoring," *Sensors*, vol. 18, no. 10, p. 3243, Sep. 2018.
[6] R. M. Allen and D. Melgar, "Earthquake early warning: Advances, scientific challenges, and societal needs," *Annu. Rev. Earth Planet. Sci.*, vol. 47, no. 1, pp. 361–388, May 2019.
[7] V. J. Hodge, S. O'Keefe, M. Weeks, and A. Moulds, "Wireless sensor networks for condition monitoring in the railway industry: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 3, pp. 1088–1106, Jun. 2015.
[8] H. Wang, W. Quan, Y. Wang, and G. Miller, "Dual roadside seismic sensor for moving road vehicle detection and characterization," *Sensors*, vol. 14, no. 2, pp. 2892–2910, Feb. 2014.
[9] M. F. Duarte and Y. Hen Hu, "Vehicle classification in distributed sensor networks," *J. Parallel Distrib. Comput.*, vol. 64, no. 7, pp. 826–838, Jul. 2004.
[10] M. A. Richards, *Fundamentals of Radar Signal Processing*. New York, NY, USA: McGraw-Hill, 2005.
[11] M. Kalra, S. Kumar, and B. Das, "Moving ground target detection with seismic signal using smooth pseudo Wigner–Ville distribution," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 3896–3906, Jun. 2020.
[12] K. Bin, Y. Long, X. Tong, and J. Lin, "Ground moving target detection with seismic fractal features," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
[13] A. Paul and T. Sato, "Localization in wireless sensor networks: A survey on algorithms, measurement techniques, applications and challenges," *J. Sensor Actuator Netw.*, vol. 6, no. 4, p. 24, Oct. 2017.
[14] P. W. Boettcher, J. A. Sherman, and G. A. Shaw, "Target localization using acoustic time-difference of arrival in distributed sensor networks," *Proc. SPIE*, vol. 4741, pp. 180–191, Aug. 2002.
[15] D. Li and Y. H. Hu, "Energy-based collaborative source localization using acoustic microsensor array," *EURASIP J. Adv. Signal Process.*, vol. 2003, no. 4, pp. 1–17, Dec. 2003.
[16] M. Z. Rahman, D. Habibi, and I. Ahmad, "Source localisation in wireless sensor networks based on optimised maximum likelihood," in *Proc. Australas. Telecommun. Netw. Appl. Conf.*, Dec. 2008, pp. 235–239.
[17] Y.-A. Le Borgne, A. Nowé, N. Abughalieh, and K. Steenhaut, "Distributed regression for high-level feature extraction in wireless sensor networks," in *Proc. 7th Int. Conf. Networked Sens. Syst. (INSS)*, Jun. 2010, pp. 249–252.
[18] A. Gouda, D. Heinrich, M. Hünnefeld, I. F. Priyanta, C. Reining, and M. Roidl, "A grid-based sensor floor platform for robot localization using machine learning," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (IMTC)*, May 2023, pp. 1–6.
[19] L. Yuan, H. Chen, R. Ewing, E. P. Blasch, and J. Li, "3-D indoor positioning based on passive radio frequency signal strength distribution," *IEEE Internet Things J.*, vol. 10, no. 15, pp. 13933–13944, Aug. 2023.
[20] S. M. Taheri and H. Nosrati, "Acoustic signature identification using distributed diffusion adaptive networks," in *Proc. 9th Int. Symp. Commun. Syst., Netw. Digit. Sign (CSNDSP)*, Jul. 2014, pp. 943–948.
[21] A. Zwartjes, P. Havinga, G. Smit, and J. Hurink, "QUEST: Eliminating online supervised learning for efficient classification algorithms," *Sensors*, vol. 16, no. 10, p. 1629, Oct. 2016.
[22] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–9.
[23] T. J. S. Chowdhury, C. Elkin, V. Devabhaktuni, D. B. Rawat, and J. Oluoch, "Advances on localization techniques for wireless sensor networks: A survey," *Comput. Netw.*, vol. 110, pp. 284–305, Dec. 2016.

[24] J. Luomala and I. Hakala, "Adaptive range-based localization algorithm based on trilateration and reference node selection for outdoor wireless sensor networks," *Comput. Netw.*, vol. 210, Jun. 2022, Art. no. 108865.
[25] E. Köse and A. K. Hocaoglu, "A new spectral estimation-based feature extraction method for vehicle classification in distributed sensor networks," *TURKISH J. Electr. Eng. Comput. Sci.*, vol. 27, pp. 1120–1131, Mar. 2019.
[26] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.
[27] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, 2013, pp. 3–11.
[28] I. Goodfellow, Y. Bengio, and A. Courville, "Softmax units for multinoulli output distributions," in *Deep Learning*, vol. 1, no. 2. Cambridge, MA, USA: MIT Press, 2016, pp. 180–184.
[29] P. J. Huber, "Robust estimation of a location parameter," in *Breakthroughs in Statistics*, S. Kotz and N. L. Johnson, Eds. New York, NY, USA: Springer, 1992, pp. 492–518.
[30] J. Brownlee, *Deep Learning With Python: Develop Deep Learning Models on Theano and Tensorflow Using Keras*. Machine Learning Mastery, 2016. [Online]. Available: https://machinelearningmastery.com/deeplearning-with-python/
[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[32] L. N. Smith, "Cyclical learning rates for training neural networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Mar. 2017, pp. 464–472.
[33] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, vol. 22, no. 3, pp. 400–407, Sep. 1951.
[34] M. Duarte and Y.-H. Hu, "Distance-based decision fusion in a distributed wireless sensor network," *Telecommun. Syst.*, vol. 26, nos. 2–4, pp. 339–350, Jun. 2004.
[35] N. Sindhu, S. Empran, M. Uttarakumari, and S. D. Badiger, "Target classification and tracking using symbolic dynamic filtering," in *Proc. 9th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2018, pp. 1–6.
[36] K. Lee, B. S. Riggan, and S. S. Bhattacharyya, "A joint target localization and classification framework for sensor networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 3076–3080.
[37] S. M. Nabritt, T. Damarla, and G. Chatters, "Personnel and vehicle data collection at Aberdeen proving ground (APG) and its distribution for research," U.S. Army Res. Lab., Adelphi, MD, USA, Tech. Rep. ARL-MR-0909, 2015.
[38] P. W. Boettcher, G. A. Shaw, and J. Sherman, "A distributed time-difference of arrival algorithm for acoustic bearing estimation," in *Proc. Int. Conf. Inf. Fusion*, vol. 1, 2002, Paper TuC32.
[39] (2020). *AAV7A1 RAM/RS—BAE Systems*. BAE Systems. [Online]. Available: https://www.baesystems.com/en-media/uploadFile/20210406061804/1434593116846.pdf
[40] *Chassis, and Winches for Tractor Truck M26. US War Department*, document TM 9-1767C, Body, 1944.
[41] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[42] I. Martinez-Alpiste, G. Golcarenarenji, Q. Wang, and J. M. Alcaraz-Calero, "Smartphone-based real-time object recognition architecture for portable and constrained systems," *J. Real-Time Image Process.*, vol. 19, no. 1, pp. 103–115, Feb. 2022.

ERDEM KÖSE received the B.E. degree in electronics engineering from Gebze Technical University, Kocaeli, Turkey, in 2014, and the M.E. degree in electronics engineering from İstanbul Technical University, İstanbul, Turkey, in 2017. He is currently pursuing the Ph.D. degree with Gebze Technical University. He is also with Qualcomm as a Senior Engineer. His current research interests include developing target detection and localization systems for border surveillance.

ALİ KÖKSAL HOCAOĞLU received the Ph.D. degree in electrical engineering from the University of Missouri-Columbia, USA, in 2000. Between 2000 and 2003, he was a Postdoctoral Research Associate with the University of Missouri-Columbia and the University of Florida, Gainesville, USA. From 2003 to 2012, he was a Chief Researcher with the Marmara Research Center. He is currently with the Department of Electronics Engineering, Gebze Technical University, Turkey. His current research interests include developing multi-sensor target detection and tracking systems for border surveillance using passive sensors.