Delft University of Technology

Learning-based NLOS Detection and Uncertainty Prediction of GNSS Observations with Transformer-Enhanced LSTM Network

Zhang, Haoming ; Wang, Zhanxin ; Vallery, H.

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Learning-based NLOS Detection and Uncertainty Prediction of GNSS Observations with Transformer-Enhanced LSTM Network

Haoming Zhang, Zhanxin Wang and Heike Vallery

Abstract— The global navigation satellite systems (GNSS) play a vital role in transport systems for accurate and consistent vehicle localization. However, GNSS observations can be distorted due to multipath effects and non-line-of-sight (NLOS) receptions in challenging environments such as urban canyons. In such cases, traditional methods to classify and exclude faulty GNSS observations may fail, leading to unreliable state estimation and unsafe system operations. This work proposes a deep-learning-based method to detect NLOS receptions and predict GNSS pseudorange errors by analyzing GNSS observations as a spatio-temporal modeling problem. Compared to previous works, we construct a transformer-like attention mechanism to enhance the long short-term memory (LSTM) networks, improving model performance and generalization. For the training and evaluation of the proposed network, we used labeled datasets from the cities of Hong Kong and Aachen. We also introduce a dataset generation process to label the GNSS observations using lidar maps. In experimental studies, we compare the proposed network with a deep-learning-based model and classical machine-learning models. Furthermore, we conduct ablation studies of our network components and integrate the NLOS detection with data out-of-distribution in a state estimator. As a result, our network presents improved precision and recall ratios compared to other models. Additionally, we show that the proposed method avoids trajectory divergence in real-world vehicle localization by classifying and excluding NLOS observations.

I. INTRODUCTION

On the way towards accurate and reliable localization of transport systems in outdoor and large-scale environments, global navigation satellite systems (GNSS) are widely used as standalone positioning solutions or sensor observations in advanced state estimation algorithms [1]. In the configuration of Real-Time Kinematic (RTK) with GNSS-correction data from a local ground-based augmentation service (GBAS), centimeter accuracy can be achieved [2]. However, the performance of GNSS-based localization approaches is highly dependent on sufficient satellite visibility and low environmental interference. These requirements are frequently violated in complex environments such as urban canyons and bridge-rich areas [3]. In such environments, infrastructure objects can easily block GNSS measurements, resulting in non-line-of-sight (NLOS) receptions and multipath effects in the received GNSS observations [4]. Compared to atmospheric

influences or satellite orbit bias, which can be corrected by the GNSS correction and ephemeris data, NLOS signals and multipath delays cannot be adequately modeled by statistical methods or eliminated using reference measurements from an external data link [5]. Therefore, consistent and robust vehicle localization in urban areas remains a challenge.

Many earlier works rely predominantly on statistical error modeling or consistency checking within the obtained GNSS observations to exclude strongly corrupted GNSS measurements for vehicle positioning. In [6] and [7], multipath errors are indirectly modeled using signal strength measures such as signal-to-noise ratio (SNR) and carrier-to-noise ratio (C/N_0). However, such methods do not scale well in environments where multipath reflections with strong signal strength exist [5]. In contrast, the receiver autonomous integrity monitoring (RAIM) approach detects GNSS measurement faults by checking the consistency of received redundant GNSS observations. After computing a protection level that yields a pre-defined integrity risk and alarm limit, the RAIM indicates faulty measurements by utilizing online statistical tests [8]. Although RAIM has become the prime technology in current GNSS solutions for fault isolation, this concept still assumes Gaussian noise distributions of GNSS observations. Moreover, it requires sufficient measurement redundancy, which cannot be fulfilled in urban areas and therefore leads to conservative fault classifications [3]. Other prominent works focus on state estimation algorithms that utilize robust error functions, such as m-estimators, or integrate other sensors to mitigate the effect of degraded GNSS measurements. In [9] and [10], the authors have proposed a tightly coupled fusion of inertial measurement alongside corrupted GNSS observations based on factor graph optimization with robust noise modeling using m-estimators and have shown superior performance against classic Bayesian filters. The principle of Graduated-Non-Convexity instead of traditional m-estimators can further mitigate the impact of strongly corrupted GNSS observations in urban areas [11]. However, these approaches cannot explicitly exclude faulty GNSS measurement and thus do not yet guarantee stability and robustness.

Another group of approaches employs prior environmental information, such as 3D city models or fish-eye images, to acquire a satellite-blocking mask in the sky plot that is conducted for further visibility argumentation [12]–[16]. Also, lidar sensing, which is usually used as an optical vehicle odometer, has been explored for GNSS fault exclusion. The generated 3D point cloud map can provide a more detailed and effective satellite-blocking mask considering local objects such as trees and large vehicles. This mask

can be used for GNSS fault exclusion in state estimators, [17, 18]. However, these approaches rely on well-calibrated third-party sensors or prior information and cannot always guarantee real-time efficiency for fault detection.

Recently, learning-based methods have drawn great attention for NLOS classification and multipath error prediction [19]. Pre-trained models can be inferred online for GNSS fault detection after intensive data augmentation and offline training. Some works have utilized classical machine learning methods such as fuzzy logic [20] or support vector machines [21, 22] to classify faulty GNSS measurements, which achieve impressive classification accuracy in test data. Considering the time-correlated and strongly nonlinear properties of faulty GNSS observations, deep-learning-based (DL) methods can extract semantic features and acquire better generalization outcomes using pre-trained models [23]–[25]. Among all of these works, Zhang et al. proposed a particular network architecture containing fully-connected layers and long short-term memory (LSTM) networks for satellite visibility classification and pseudorange error prediction [26]. In [27], the pre-trained network in [26] has been utilized in a state estimator to exclude the NLOS observations and shows promising accuracy improvement. Although LSTM networks are considered a feasible method to model sequential data [28]. Therefore, they outperform other approaches for NLOS detection [25]. However, LSTM networks are limited in applications with multivariate features and contextual information [29]. To overcome this problem, the self-attention mechanism introduced as transformers can be utilized to improve performance in sequential data modeling [30].

Inspired by [26] and [28], we consider the GNSS observations as time-series data and propose a transformer-enhanced LSTM network for NLOS classification and pseudorange error prediction. We follow the noise characteristics of the GNSS observations in complex environments and consider both temporal and spatial information in the feature extraction process. To train the network, we use the data set in [26] and introduce a data generation process, which is used to label the GNSS data from our measurement campaign in the city of Aachen. We train the proposed network alongside a baseline LSTM network in [26] and two classical machine learning models: support vector machine (SVM) and extreme gradient boosting (XGBoost). All models are evaluated with test and unseen (out-of-distribution) data. We also conduct ablation studies on our network components and present comprehensive discussions on model performance, data set quality, and the importance of features. Furthermore, we employ the pre-trained models on unlabeled data from our measurement campaign in the city of Düsseldorf to infer NLOS classifications, which are integrated into a state estimator.

For the benefit of the research community, we release our code and Aachen dataset in `github`[1].

II. DATASET GENERATION

A. Data Collection

For the training and testing of the learning-based algorithm, we collected raw data in a measurement campaign in Aachen and proposed a data labeling process. The measurement campaign contains a route with a length of $17\,\text{km}$ in different urban areas. The dataset consists of raw measurements from an inertial measurement unit (IMU) at $100\,\text{Hz}$, lidar point clouds at $10\,\text{Hz}$, and GNSS measurements at $10\,\text{Hz}$ using a NovAtel PwrPak7 receiver with RTCMv3 correction data received from the German GBAS server[2].

B. Label Generation

Fig. 1 shows our label generation process. To obtain the pseudorange residuals and satellite blocking masks, we use a multi-sensor fusion state estimator and a GNSS preprocessor from our previous work [31]. The raw GNSS measurements with the ephemeris data and RTCMv3 correction data are processed in the GNSS preprocessor to eliminate the satellite orbit bias and the atmospheric delays. After GNSS pre-processing, we feed the GNSS observations with the IMU measurements and lidar point clouds into the state estimator, where the pseudorange residuals are calculated and 3D lidar maps are generated. The 3D lidar maps are used to mark the satellite blockages in the sky plots, which are further conducted for label generation. Fig. 2 shows the masked sky plots. While generating labels, we query the neighbored elevation and azimuth angles at each satellite and infer the satellite with NLOS reception if the elevation/azimuth of the satellite is masked according to and the pseudorange residual is above a predefined threshold.
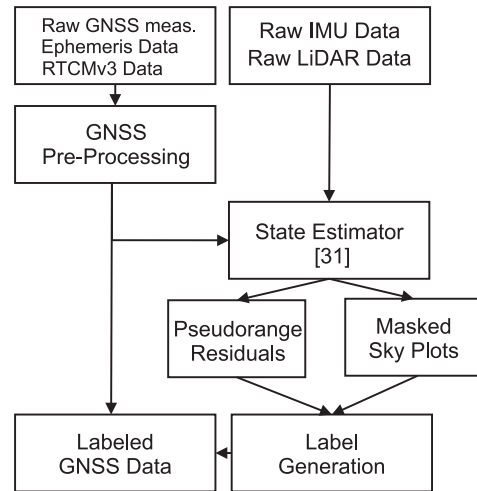

Fig. 1: Proposed labeling framework

C. GNSS Feature Selection

Following the previous work [26], we select the same five features from the pre-processed GNSS observations for the NLOS classification and pseudorange error prediction.
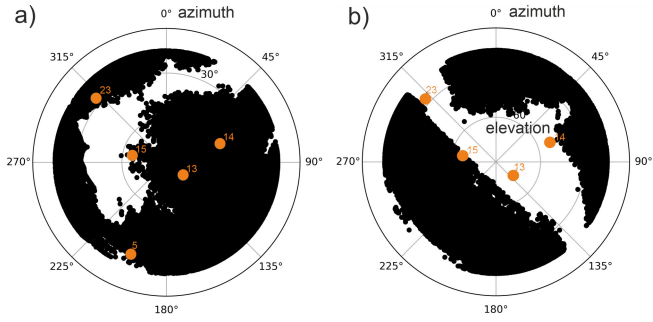
Fig. 2: Sky plots from two scenes: a) urban tree-rich area, b) urban. The received GPS satellites are marked in orange.

1) Elevation Angle (El)*:* The satellite elevation angle El provides an adequate quality measure of the satellite visibility. Satellites with higher elevation angles are generally less affected by the infrastructure objects. In classical GNSS positioning approaches, satellites with low elevation angles are excluded from state estimation algorithms. The elevation angle El_k of k-th satellite can be calculated as

$$\text{El}^k = \arcsin(h^{\text{sat},k}/r^k), \tag{1}$$

where the satellite's height with respect to the vehicle in the East-North-Up (ENU) frame is represented as $h^{\text{sat},k}$. The variable r^k denotes the geometric range from the vehicle to the k-th satellite.

2) Azimuth Angle (Az)*:* The satellite azimuth angle represents the horizontal geometric relation between the satellite and the vehicle. As described in [26], azimuth angles allow deducing the satellite visibility if multiple satellites with similar azimuth angles but different elevation angles are received. This feature is expected to provide more information on the geometric context in learning-based methods. We calculate the azimuth angle with

$$\text{Az}^k = \arctan(x^{\text{sat},k}/y^{\text{sat},k}), \tag{2}$$

where the symbols $x^{\text{sat},k}$ and $y^{\text{sat},k}$ are the positions of k-th satellite relative to vehicle's position in the ENU-frame.

3) Carrier-to-Noise-Ratio (C/N_0)*:* The Carrier-to-Noise Ratio C/N_0 measures the satellite signal strength and characterizes the satellite signal quality. Generally, the GNSS observations used for positioning are expected to have higher values C/N_0. In urban scenarios, C/N_0 can also effectively represent interference due to NLOS receptions and multipath effects [5].

4) Least-Square Pseudorange Residual $(\sigma_{\text{LS}}^{\text{sat}})$*:* In many GNSS standalone solutions, vehicle position can be calculated with an iterative least-square (LS) method using GNSS observations. We use this positioning solution to calculate pseudorange residuals, denoted as $\sigma_{\text{LS}}^{\text{sat},k}$. Unlike the pseudorange residuals calculated in the state estimator by combining multiple sensors, the residual $\sigma_{\text{LS}}^{\text{sat},k}$ is highly related to the quality of GNSS observations and represents the scale of pseudorange errors. We calculate the LS pseudorange residual by

$$\sigma_{\text{LS}}^{\text{sat},k} = \hat{\rho}_k - \|\hat{\boldsymbol{x}}^r - \boldsymbol{x}^{\text{sat},k}\|_2, \tag{3}$$

where the range $\hat{\rho}_k$ is the measured pseudorange, which is corrected in the GNSS pre-processing process to eliminate the atmospheric delays. The satellite position and the estimated vehicle position are represented as $\boldsymbol{x}^{\text{sat},k}$ and $\hat{\boldsymbol{x}}^r$, respectively.

5) Root-Sum-Squares of Pseudorange Residuals (RSS)*:* As introduced in [26], the root-sum-square of the pseudorange residuals encodes the environmental condition by summing up multiple pseudorange residuals in a certain time window. If the vehicle is moving in urban areas in $t = 1...T$, larger RSS should be expected. We employ the RSS, shown in

$$\text{RSS} = \sqrt{\sum_{t=1}^{T} (\sigma_{\text{LS},t}^{\text{sat},k})^2} \tag{4}$$

as the last feature in the feature vector.

III. PROPOSED NETWORK

A. Problem Formulation

The pseudorange observations are commonly used to represent the vehicle position in GNSS-based localization approaches, as shown in

$$\rho_t^k = r_t^{\text{sat},k} + c(\delta t_r - \delta t^{\text{sat},k}) + T(t) + I(t) + M + \epsilon_k, \tag{5}$$

$$\hat{\rho}_t^k = r_t^{\text{sat},k} + M + \epsilon_k, \tag{6}$$

where the range $r_t^{\text{sat},k} = \|\boldsymbol{x}_t^r - \boldsymbol{x}_t^{\text{sat},k}\|$ is the geometric range between the vehicle and k-th satellite. The receiver clock error δt_r and the satellite clock error $\delta t^{\text{sat},k}$, multiplied by the light speed c, represent the clock bias. The variables $T(t)$ and $I(t)$ are tropospheric and ionospheric delays, respectively.

In this work, we assume that the receiver clock bias is estimated in the state estimator and therefore is subtracted from (5). With the ephemeris and differential correction data, the satellite clock bias and atmospheric delays can be eliminated in a pre-processing process to acquire the corrected pseudorange $\hat{\rho}_t^k$. We denote the multipath delay and the observation noise using M and ϵ_k, respectively. These delays vary particularly in urban areas and become complex and time-related due to NLOS receptions and multipath effects [32].

Therefore, we consider the NLOS detection and error prediction of pseudorange $\hat{\rho}_t^k$ as a spatio-temporal modeling problem

$$[\mathcal{V}^k, \ \hat{\epsilon}^k] = \text{NN}\left(\left\{\text{El}_\tau^k, \ \text{Az}_\tau^k, \ \text{C/N}_{0\tau}^k, \ \sigma_{\text{LS},\tau}^k, \ \text{RSS}^k\right\}_{\tau=1}^{T}\right), \tag{7}$$

which predicts the satellite visibility $\mathcal{V}^k$ and pseudorange error $\hat{\epsilon}^k$ using a neural network given all selected features in a time series of T seconds.
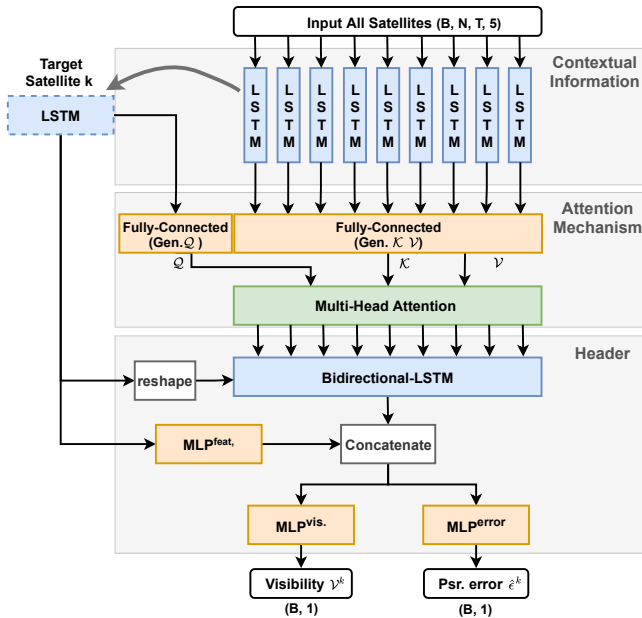
Fig. 3: Proposed Neural Network Architecture.

B. Neural Network Architecture

Considering the noise characteristics of GNSS observations in urban areas, we propose a neural network containing three properties: 1. capturing the temporal variations of selected features, 2. evaluating spatial contextual information from other satellites, and 3. providing effective attention for discriminative features. The proposed neural network architecture is presented in Fig. 3.

Assuming that N satellites can be received at maximum, we formulate the input tensor as a T second time series of feature vectors from each satellite in a mini-batch B. To extract high-level features that contain temporal information from the received observations, we process the time series of each satellite with LSTM networks individually. The LSTM output features of all satellites are forwarded to a fully connected layer to generate contextual key $\mathcal{K}$ and value $\mathcal{V}$ features equivalent to transformer networks. Meanwhile, we detach the LSTM output features as input for three network components to obtain predictions for each target satellite. Above all, we employ the same fully-connected layer to generate the query features $\mathcal{Q}$ for the attention mechanism. The features $\mathcal{K}$, $\mathcal{V}$, and $\mathcal{Q}$ are evaluated in the multi-head attention sub-network, which is designed to propose spatial discriminative features from contextual information. In the header stack, we employ a bi-directional LSTM (Bi-LSTM) by initializing its hidden states with the output features from the foregoing LSTM network of the target satellite. The output features of the attention mechanism are processed in the Bi-LSTM, which is expected to extract the temporal information of the discriminative features and mitigate the effect of satellite ordering in the input tensor. We concatenate the output of the Bi-LSTM network with LSTM features of each target satellite to obtain high-level semantic features for classification and regression problems, where a binary

TABLE I: Details of Datasets. The average and maximal number of the received satellite are denoted with $n_{\text{avg}}^{\text{sat}}$ and $n_{\text{max}}^{\text{sat}}$. The maximal pseudorange error $\sigma_{\text{max}}^{\rho}$ is inferred with ground-truth positioning. We present the ratio of LOS and NLOS labels in the dataset with R^{LOS} and R^{NLOS}.

Dataset	Duration s	$n_{\text{avg}}^{\text{sat}}$	$n_{\text{max}}^{\text{sat}}$	$\sigma_{\text{max}}^{\rho}$ m	R^{LOS} %	R^{NLOS} %
HK [26]	10573	17	25	1058.13	60.62	39.38
AC	2366	6	8	640.32	81.13	18.87
DUS[3]	810	7	11	355.74	-	-

visibility indicator $\mathcal{V}^k$ and a predicted pseudorange error $\hat{\epsilon}^k$ of the target satellite k are inferred, as shown in Fig., 3.

Because our input tensor does not contain high-dimensional features, LSTM networks are constructed with two hidden layers. We implement the MLPs with three hidden layers and use ReLU as the activation function. For more details, see the code introduced in Sec. I.

C. Implementation and Training

1) Training and Test Data: In this work, we use the Hong Kong (HK) dataset [26] and our labeled dataset in Aachen (AC) for training and testing. All dataset details are shown in Tab. I. Compared to the HK dataset, where more NLOS observations can be received due to the high urbanization rate, the AC dataset only contains 18.87% of NLOS receptions, presenting an unbalanced dataset. To evaluate the model generalization with data out-of-distribution, we use unlabeled data recorded in Düsseldorf (DUS) and predict the satellite visibility. This prediction is validated using a state estimation algorithm.

2) Training Protocol: We implement all models using `PyTorch`[4]. deep-learning (DL) models are trained with the `Adam` optimizer [33] by utilizing a multistep learning rate scheduler. In the implementation, we use the MSE loss for the NLOS classification and L1 loss in the regression problem to predict the pseudorange error.

We employ the basic implementation in `Sklearn`[5] for all classical machine-learning models. The SVM classifier was trained using the RBF kernel with the penalty parameter $C = 1.0$. We implement XGBoost with 200 decision trees and a learning rate $\alpha = 1.0$ for the classification task.

The machine learning models are only trained for NLOS classification, whereas the DL models conduct two loss functions both for NLOS classification and pseudo-range error prediction. We train all models with datasets from Aachen and Hong Kong and consider testing procedures using data from the same and different datasets, with the aim of validating the generalization of the model.

IV. EXPERIMENTAL RESULTS

A. General Performance Metrics for NLOS Classification

1) Data In-Distribution: To evaluate the superiority of the proposed network, we also train different classical machine-

---

[3]Only used for prediction.
[4]https://pytorch.org/
[5]https://scikit-learn.org/

TABLE II: Performance Metrics using HK Dataset.

Model		Precision	Recall	F1-Score	Acc.
SVM	LOS	0.87	0.86	0.86	0.82
	NLOS	0.72	0.74	0.73	
XGboost	LOS	0.90	0.88	0.89	0.85
	NLOS	0.77	0.80	0.78	
DL [26]	LOS	0.80	0.96	0.87	0.81
	NLOS	0.85	0.51	0.64	
DL (ours)	LOS	0.90	0.91	0.90	0.87
	NLOS	0.80	0.79	0.80	

TABLE III: Performance Metrics using AC Dataset.

Model		Precision	Recall	F1-Score	Acc.
SVM	LOS	0.80	1.00	0.89	0.80
	NLOS	0.39	0.01	0.01	
SVM*	LOS	0.91	0.72	0.80	0.72
	NLOS	0.38	0.70	0.49	
XGboost	LOS	0.86	0.88	0.87	0.79
	NLOS	0.47	0.42	0.45	
DL [26]	LOS	0.86	0.83	0.84	0.76
	NLOS	0.40	0.47	0.43	
DL (ours)	LOS	0.86	0.90	0.88	0.81
	NLOS	0.51	0.42	0.46	

TABLE IV: Cross-Dataset Performance Metrics: The models are trained using the HK dataset and tested with the AC dataset.

Model		Precision	Recall	F1-Score	Acc.
SVM	LOS	0.88	0.82	0.85	0.77
	NLOS	0.42	0.54	0.47	
XGboost	LOS	0.87	0.74	0.80	0.70
	NLOS	0.32	0.54	0.40	
DL [26]	LOS	0.86	0.81	0.84	0.74
	NLOS	0.35	0.45	0.39	
DL (ours)	LOS	0.89	0.79	0.84	0.75
	NLOS	0.40	0.60	0.48	

learning approaches and the network proposed in [26] in the experimental study. Tab. II and Tab. III show the general performance metrics of all models using the same datasets for training and testing.

Because the HK dataset presents well-balanced NLOS observations compared to the AC dataset, as shown in Tab. I, all models can achieve fair performance for both LOS and NLOS classifications. Among all, the classical machine-learning models also present high precision and recall ratios and slightly outperform the model introduced in [26]. Our implementation of the model [26] can reproduce similar performance metrics and shows higher precision due to different training protocols. Compared to [26], our proposed model with the attention mechanism outperforms all other models and presents both high precision and recall ratios.

We employ the same training and testing configurations using the AC dataset that contains major LOS observations. The performance metrics are shown in Tab. III. In contrast to the LOS observations, which can be classified with high accuracy and recall ratios, the performance for the NLOS classification of all models degrades dramatically. Classical models like SVM are more sensitive to imbalanced datasets than DL models. This has also been discussed in [34]. To validate this conclusion, we retrain the SVM using a modified AC dataset containing equally sampled NLOS and LOS data, denoted as SVM*. It can be shown from this experiment that the SVM trained with a balanced dataset shows improved performance compared to other models. On the contrary, DL models present expected learning performance even when trained on unbalanced datasets.

2) Data Out-Of-Distribution: Furthermore, we conducted the experiment to evaluate the performance of pre-trained models on data out-of-distribution. As the HK dataset contains a fair distribution of the NLOS and LOS observations, we train all models using the HK dataset and test them with the AC dataset. The results are shown in Tab. IV. In

this experiment, the SVM and the proposed DL slightly outperform other models and show similar performance compared to each other. However, all models cannot show fair performance in the NLOS classification compared to Tab. III, where the classification precision and especially the recall ratio are more elevated.

Discussion: For the NLOS classification task, we show that all learning-based models perform well by training with a well-balanced dataset. Even classical models present both high-precision and recall ratios. One possible reason for this result is that these models fit overwhelmingly on some features, which are highly related to NLOS receptions. In contrast, the contextual features that indirectly infer GNSS faults are discounted. If the classical models are inferred with data out-of-distribution, the recall ratio drops dramatically. Thus, we expect that the DL networks can present more generalized models. When comparing the DL models, our network is optimized with an attention mechanism that effectively captures the spatial and temporal information from the GNSS observations, showing better performance. However, the proposed method presents higher model complexity, which requires a more thorough training procedure and penalizes runtime efficiency. Considering the dataset quality, we emphasize the importance of label balance in the datasets, which plays a vital role in the performance of learning-based methods. We also show the performance degradation of all models while testing using data out-of-distribution. One probable reason can be the inherent quantity difference between the two datasets. For instance, we calculate the maximum, minimum, and average C/N_0 in two datasets used for training, shown in Tab. V. Suppose that the model has been pre-trained using a dataset with data bias. In that case, the model will present a distribution shift, leading to performance degradation, as discussed in [35].

TABLE V: Analysis of C/N_0 in Two Datasets.

Dataset	max. C/N_0 dB-Hz	min. C/N_0 dB-Hz	avg. C/N_0 dB-Hz
HK [26]	49.0	7.0	34.0
AC	57.0	22.0	49.0

B. Performance of Pseudorange Error Prediction

We only evaluated the baseline DL models in the pseudorange error prediction task. The predicted pseudorange
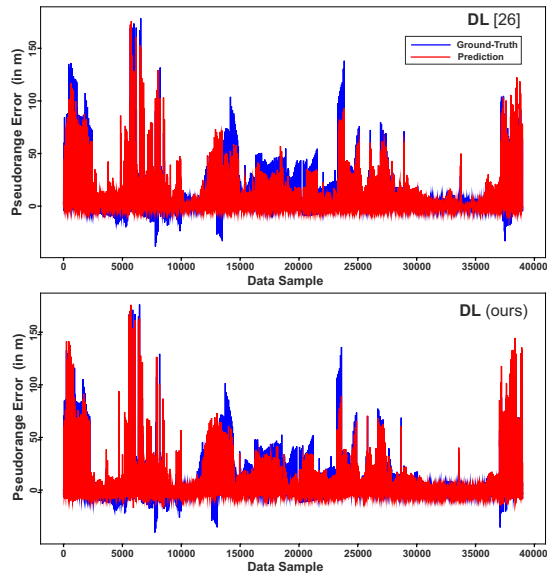
Fig. 4: Predicted and ground-truth pseudorange errors with the model [26] and our network.

Fig. 5: Feature Permutation Importance of different models.

TABLE VI: Performance metrics during ablation studies of the proposed method.

Model		Precision	Recall	F1-Score	Acc.
w/o Att. Mecha.	LOS	0.83	0.91	0.87	0.82
	NLOS	0.76	0.63	0.69	
w/o BiLSTM	LOS	0.81	0.91	0.86	0.80
	NLOS	0.76	0.57	0.65	

errors and ground-truth are illustrated in Fig. 4, presenting a qualitative performance view. Compared to positive pseudorange errors, both models do not perform well in predicting negative and large pseudorange errors, which have a non-negligible impact on consistent state estimation. The average prediction error when testing the proposed model amounts to $5.68\,\mathrm{m}$, which is lower than $6.53\,\mathrm{m}$ from the model in [26].

Discussion: Even though the proposed network presents a higher prediction accuracy, we indicate that the current pre-trained DL models are not able to infer plausible pseudorange error predictions. Because there are multiple interference sources for the pseudorange, faulty pseudoranges present highly complex noise distributions. Thus, the most probable reason is the insufficient data samples of faulty pseudoranges with large errors. This could be improved if the network structure for the regression task is further optimized and more faulty GNSS observations can be sampled for training.

C. Ablation Studies

Besides general performance metrics, we propose ablation studies on feature importance and different components of our DL model.

1) Feature Importance: Although the classical models can achieve the same high accuracy as the DL models when comparing the results in Tab. II and Tab. III, the DL models show better generalization considering unbalanced or unseen data. One reason to support this hypothesis can be speculated with the feature permutation importance, as shown in Fig. 5. While classical models significantly rely on a few features, such as elevation angle and $\mathrm{C/N_0}$, DL models tend to rate all features equally. This behavior should be expected in this learning problem because features such as azimuth angle and the Root-Sum-Squares (RSS) provide crucial contextual information about signal blockage in urban
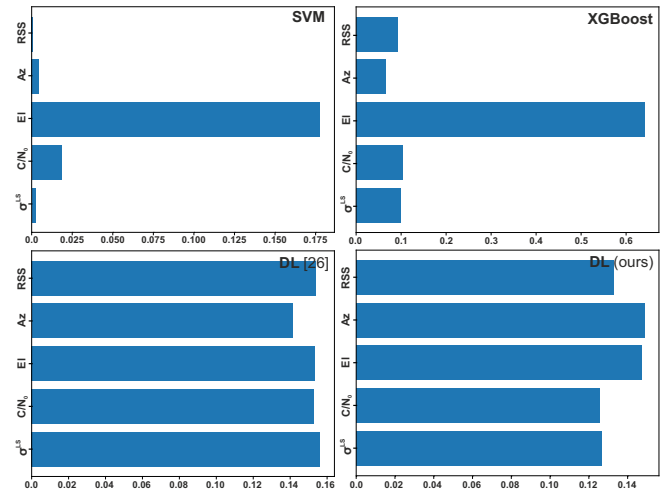
areas. A fair weighting of all features can contribute more to model generalization and, thus, avoid the overfitting effect.

2) Model Components: To prove our hypothesis in the network design, we evaluated different components of our model and discuss the effect of integrating the attention mechanism and the Bi-LSTM network. For this experiment, the HK dataset and identical training settings as in Tab. II were employed to propose the results in Tab. VI.

As introduced in Sec. III-B, handling spatial information using the attention mechanism benefits both model precision and recall ratio for the NLOS classification, even when the dataset contains more LOS observations. Replacing the Bi-LSTM with a normal LSTM network penalizes the recall ratio for NLOS detection. This phenomenon confirms the superiority of the Bi-LSTM, as introduced in [36], that the Bi-LSTM shows a large model capacity by preserving information from both the past and future, especially when contextual information is non-negligible.

D. Vehicle Localization with GNSS Fault Exclusion

In this experiment, we employ the pre-trained models to infer the NLOS receptions of GPS-L1 and Galileo-E1 observations obtained using a NovAtel PwrPak7 receiver in a measurement campaign in Düsseldorf. The GNSS observations are not used for training and, thus, present data out-of-distribution. By employing the pre-trained model using HK and AC datasets, the predicted ratio of NLOS observations of each model is shown in Tab.,VII.

We integrate the NLOS detections of all models in a graph-optimization-based state estimation algorithm that fuses the GNSS pseudorange and Doppler shift with IMU measure-
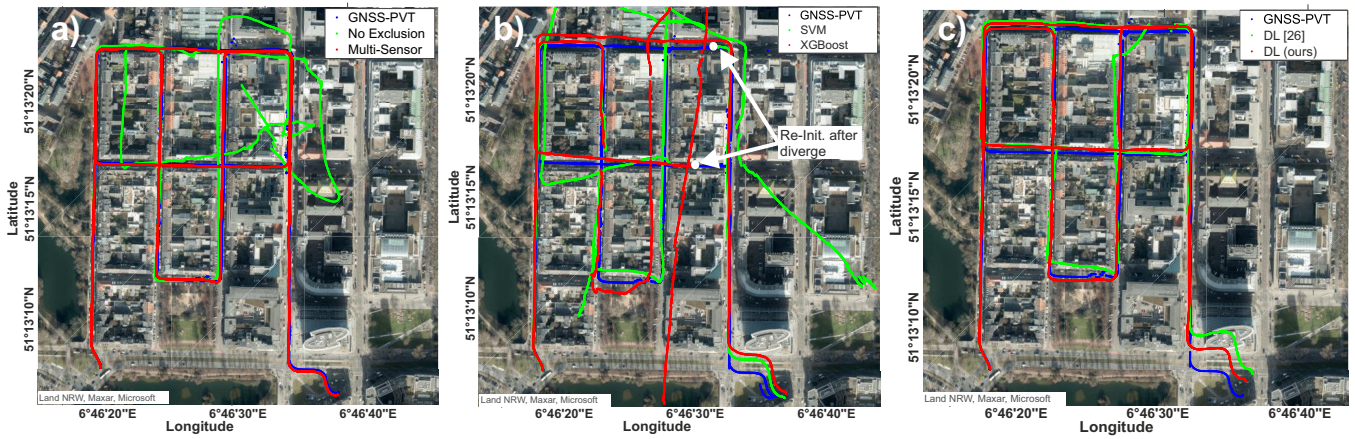
915

Fig. 6: Vehicle Localization in Urban Area in Düsseldorf with NLOS Exclusion. The GNSS Positioning-Velocity-Timing (PVT) solution is shown in blue. a) presents the estimated trajectory without NLOS exclusion and the solution by fusing GNSS observations and lidar odometry in a tight coupling. b) and c) illustrate the estimated trajectories with NLOS exclusion using different learning models.

TABLE VII: Result of NLOS classification with data out-of-distribution.

Model	$\hat{R}^{LOS}$ %	$\hat{R}^{NLOS}$ %
SVM	64.83	35.17
XGBoost	58.68	41.32
DL [26]	76.74	23.26
DL (ours)	74.76	25.24

ments in a tight coupling [31]. Here, we consider the dual-constellation of GPS and Galileo measurements, which are preprocessed with RTCM correction data. The pseudorange observations, which are classified as NLOS receptions, are excluded for integration into the graph.

As shown in Fig. 6 a), the GNSS observations in challenging environments are strongly corrupted. Thus, the state estimator integrates observations with inconsistent sensor measurement models, leading to a divergence in trajectory estimation. This problem can be alleviated if the NLOS observations are effectively excluded, illustrated in Fig. 6 c).

Compared to DL models, more NLOS observations are identified using SVM and XGBoost. This reduces constraints that are factorized from the GNSS observations in the optimization problem, leading to dramatic trajectory drifting and divergence, as shown in Fig.,6 b). Our hypothesis of this problem can be referred to Sec. IV-C.1. The classical models generally count on a few features. They are not capable of extracting and evaluating high-level features, resulting in overconfident feature weighting, and thus, proposing more False-Positive NLOS classifications. The DL models identify more LOS observations, which contributes to more sensor observations integrated for state estimation. On this basis, consistent trajectories are estimated if NLOS observations can be effectively classified, as demonstrated in Fig.,6 c).

Discussion: In this experiment, our network infers more NLOS receptions compared to the model in [26]. The vehicle trajectory can be consistently estimated by excluding the NLOS observations classified with our network, which indirectly confirms that the proposed network can infer more true-positive NLOS detections. However, the NLOS exclusion can improve the consistency of the trajectory estimation, but accuracy cannot yet be guaranteed. This result can be associated with the insufficient number of GNSS observations that survive after the NLOS exclusion in our measurement campaign. Thus, since the NLOS detector contributes a majorly to trajectory consistency, integrating other sensors into the state estimator still presents a feasible way to improve the precision of estimated trajectories in challenging environments.

V. CONCLUSION AND FUTURE WORK

In this work, we proposed a transformer-enhanced LSTM network to detect NLOS receptions and predict pseudorange errors of GNSS observations. The proposed network was evaluated with a baseline LSTM network and classical machine learning models using two datasets from Hong Kong and Aachen. Besides, we introduced a dataset generation process using a multi-sensor state estimator. We conducted experimental studies of the pre-trained models both for data in-distribution and out-of-distribution. The results show that the deep-learning-based methods proposed more generalized models, which present considerable inference performance for the NLOS classification. Moreover, the DL models are less prone to overfitting compared to classical machine learning models. Among all models, the proposed network, enhanced by the attention mechanism, captures both temporal and spatial information with all satellite observations as contextual information and has achieved the best performance even with data out-of-distribution. Furthermore, our model shows effective NLOS detection in real-world vehicle localization, where the consistent trajectory can be estimated. For future work, more balanced datasets can be utilized to train the proposed network. And the input tensor can be extended with other advanced features, such as the Doppler

rate consistency. We also plan to integrate the proposed method into our online state estimator for effective NLOS rejection [31].

ACKNOWLEDGMENTS

REFERENCES

[1] P. Groves, *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems, Second Edition*. Artech House, 2013.

[2] Y. Feng and J. Wang, "GPS RTK performance characteristics and analysis," *Journal of Global Positioning Systems*, vol. 7, 06 2008.

[3] N. Zhu, J. Marais, D. Bétaille, and M. Berbineau, "GNSS position integrity in urban environments: A review of literature," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 9, pp. 2762–2778, 2018.

[4] P. Groves, "GNSS solutions: Multipath vs. NLOS signals. how does Non-Line-of-Sight reception differ from multipath interference," *Inside GNSS (Magazine)*, vol. 8, pp. 40–42, 12 2013.

[5] L.-T. Hsu, "Analysis and modeling GPS NLOS effect in highly urbanized area," *GPS Solutions*, vol. 22, 11 2017.

[6] Z. Zhang, B. Li, Y. Gao, and Y. Shen, "Real-time carrier phase multipath detection based on dual-frequency C/N0 data," *GPS Solutions*, vol. 23, 11 2018.

[7] C. Comp and P. Axelrad, "Adaptive SNR-based carrier phase multipath mitigation technique," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 1, pp. 264–276, 1998.

[8] T. Walter and P. Enge, "Weighted RAIM for precision approach," in *Proc. 8th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 1995)*, Sept. 24, 1995.

[9] W. Wen, T. Pfeifer, X. Bai, and L.-T. Hsu, "Factor graph optimization for GNSS/INS integration: A comparison with the extended Kalman filter," *NAVIGATION*, vol. 68, no. 2, pp. 315–331, 2021.

[10] H. Zhang, X. Xia, M. Nitsch, and D. Abel, "Continuous-Time factor graph optimization for trajectory smoothness of GNSS/INS navigation in temporarily gnss-denied environments," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 9115–9122, 2022.

[11] W. Wen, G. Zhang, and L.-T. Hsu, "GNSS outlier mitigation via graduated non-convexity factor graph optimization," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 1, pp. 297–310, 2022.

[12] S. Xin, J. Geng, G. Zhang, H.-F. Ng, J. Guo, and L.-T. Hsu, "3D-mapping-aided PPP-RTK aiming at deep urban canyons," *Journal of Geodesy*, vol. 96, 2022.

[13] H.-F. Ng, G. Zhang, Y. Luo, and L.-T. Hsu, "Urban positioning: 3D mapping-aided GNSS using dual-frequency pseudorange measurements from smartphones," *NAVIGATION*, vol. 68, no. 4, pp. 727–749.

[14] H.-F. Ng, L.-T. Hsu, M. J. L. Lee, J. Feng, T. Naeimi, M. Beheshti, and J.-R. Rizzo, "Real-time loosely coupled 3DMA GNSS/Doppler measurements integration using a graph optimization and its performance assessments in urban canyons of new york," *Sensors*, vol. 22, no. 17, 2022.

[15] X. Bai, W. Wen, and L.-T. Hsu, "Using sky-pointing fish-eye camera and LiDAR to aid GNSS single-point positioning in urban canyons," *IET Intelligent Transport Systems*, vol. 14, no. 8, pp. 908–914, 2020.

[16] M. J. L. Lee, S. Lee, H.-F. Ng, and L.-T. Hsu, "Skymask matching aided positioning using sky-pointing fisheye camera and 3D city models in urban canyons," *Sensors*, vol. 20, no. 17, 2020.

[17] X. Liu, W. Wen, F. Huang, H. Gao, Y. Wang, and L.-T. Hsu, "3D LiDAR aided GNSS NLOS mitigation for reliable GNSS-RTK positioning in urban canyons," 2022. [Online]. Available: https://arxiv.org/abs/2212.05477

[18] X. Liu, W. Wen, and L.-T. Hsu, "3D LiDAR aided GNSS real-time kinematic positioning via coarse-to-fine batch optimization for high accuracy mapping in dense urban canyons," *Proc. of the 35th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2022)*, 2022.

[19] L.-T. Hsu, "What are the roles of artificial intelligence and machine learning in GNSS positioning?" *Inside GNSS*, pp. 20–27, 2020.

[20] R. Sun, L.-T. Hsu, D. Xue, G. Zhang, and W. Y. Ochieng, "GPS signal reception classification using adaptive neuro-fuzzy inference system," *The Journal of Navigation*, vol. 72, no. 3, p. 685–701, 2019.

[21] L.-T. Hsu, "GNSS multipath detection using a machine learning approach," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 2017, pp. 1–6.

[22] T. Suzuki and Y. Amano, "NLOS multipath classification of GNSS signal correlation output using machine learning," *Sensors*, vol. 21, no. 7, 2021.

[23] Y. Quan, L. Lau, G. W. Roberts, X. Meng, and C. Zhang, "Convolutional neural network based multipath detection method for static and kinematic gps high precision positioning," *Remote Sensing*, vol. 10, no. 12, 2018.

[24] R. Zawislak, M. Greiff, K. Kim, K. Berntorp, S. Di Cairano, M. Konishi, K. Parsons, P. V. Orlik, and Y. Sato, "GNSS multipath detection aided by unsupervised domain adaptation," *Proc. of the 35th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2022)*, pp. 2127–2137, 2022.

[25] A. Siemuri, K. Selvan, H. Kuusniemi, P. Valisuo, and M. S. Elmusrati, "A systematic review of machine learning techniques for GNSS use cases," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 58, no. 6, pp. 5043–5077, 2022.

[26] G. Zhang, P. Xu, H. Xu, and L.-T. Hsu, "Prediction on the urban GNSS measurement uncertainty based on deep learning networks with long short-term memory," *IEEE Sensors Journal*, vol. 21, no. 18, pp. 20563–20577, 2021.

[27] X. Bai, W. Wen, G. Zhang, H.-F. Ng, and L.-T. Hsu, "GNSS outliers mitigation in urban areas using sparse estimation based on factor graph optimization," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, 2022, pp. 197–202.

[28] B. Lim, S. O. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, 2021.

[29] Y. Li, Z. Zhu, D. Kong, H. Han, and Y. Zhao, "EA-LSTM: Evolutionary attention-based lstm for time series prediction," *Knowledge-Based Systems*, vol. 181, p. 104785, 2019.

[30] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017.

[31] H. Zhang, C.-C. Chen, H. Vallery, and T. D. Barfoot, "GNSS/multisensor fusion using continuous-time factor graph optimization for robust localization," 2023. [Online]. Available: https://arxiv.org/abs/2309.11134

[32] M. Goode, S. Edwards, and P. Moore, "Time correlation in GNSS precise point positioning," *Proc. of the 26th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2013)*, pp. 1207–1214, 9 2013.

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.

[34] R. Batuwita and V. Palade, *Class Imbalance Learning Methods for Support Vector Machines*. John Wiley & Sons, Ltd, 2013, ch. 5, pp. 83–99.

[35] O. Wiles, S. Gowal, F. Stimberg, S.-A. Rebuffi, I. Ktena, K. Di Dvijotham, and A. T. Cemgil, "A fine-grained analysis on distribution shift," in *International Conference of Learning Representations*, 2022.

[36] Y.-L. He, L. Chen, Y. Gao, J.-H. Ma, Y. Xu, and Q.-X. Zhu, "Novel double-layer bidirectional LSTM network with improved attention mechanism for predicting energy consumption," *ISA Transactions*, vol. 127, pp. 350–360, 2022.