UNILoc: Unified Localization Combining Model-Based Geometry and Unsupervised Learning

N.B. When citing this work, cite the original published paper.

(article starts on next page)

UNILoc: Unified Localization Combining Model-Based Geometry and Unsupervised Learning

Yuhao Zhang*, Guangjin Pan*, Musa Furkan Keskin*, Ossi Kaltiokallio†, Mikko Valkama†, Henk Wymeersch*

*Department of Electrical Engineering, Chalmers University of Technology, Sweden

†Unit of Electrical Engineering, Tampere University, Finland

Abstract—**Accurate mobile device localization is critical for emerging 5G/6G applications such as autonomous vehicles and augmented reality. In this paper, we propose a unified localization method that integrates model-based and machine learning (ML)-based methods to reap their respective advantages by exploiting available map information. In order to avoid supervised learning, we generate training labels automatically via optimal transport (OT) by fusing geometric estimates with building layouts. Ray-tracing based simulations are carried out to demonstrate that the proposed method significantly improves positioning accuracy for both line-of-sight (LoS) users (compared to ML-based methods) and non-line-of-sight (NLoS) users (compared to model-based methods). Remarkably, the unified method is able to achieve competitive overall performance with the fully-supervised fingerprinting, while eliminating the need for cumbersome labeled data measurement and collection.**

Index Terms—**localization, machine learning, unsupervised learning, map information, optimal transport.**

I. INTRODUCTION

Accurate localization of mobile devices is critical for a wide range of applications, including emergency services, navigation, autonomous driving, and industrial internet of things (IoT) [1]–[3]. While global navigation satellite system (GNSS) provides a viable solution for rural outdoor scenarios, their effectiveness degrades significantly in dense urban environments and indoor settings due to signal blockage and multipath interference. Wireless communication systems, e.g., 5G/6G networks, present new opportunities for high-precision positioning by leveraging, for example, wide bandwidths, potentially high carrier frequency, and directional beamforming to extract fine-grained channel state information (CSI) [1]–[5].

In general, wireless localization can be broadly categorized into two main classes: 1) model-based methods that rely on channel parameter estimation and geometric relationships [2]–[5]; 2) data-driven methods that leverage machine learning (ML) to extract user position from CSI [6]–[8]. Specifically, some classical model-based methods perform well in line-of-sight (LoS) scenarios, e.g., user 1 in Fig. 1 [9]–[12], but would struggle to keep their superiority in non-line-of-sight (NLoS) scenarios[1] due to LoS blockage (e.g., user 2 in
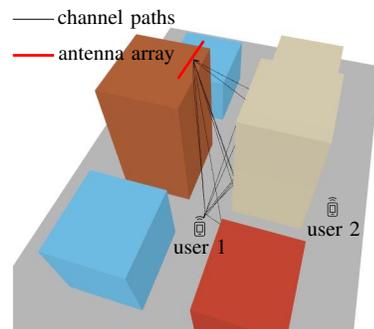

Figure 1: A street canyon scenario [17], where channel paths are shown for user 1 for example. Note that user 1 is a LoS user, while user 2 is a NLoS user as its LoS path is blocked by a building.

Fig. 1). Although more sophisticated methods that utilize NLoS paths can be adopted, the complexity (especially for channel estimation) is high and their use is usually limited to certain scenarios (e.g., relying on specular reflections) [2]–[6]. On the contrary, data-driven methods (including supervised and unsupervised learning) hold some promise for NLoS scenarios, but suffer from certain drawbacks. For example, supervised learning, e.g., fingerprinting [6]–[8], requires extensive data measurement and collection for each specific setting, which may not be practical especially for dynamic environments. On the other hand, although unsupervised channel charting can be used for localization, position estimates are highly distorted, and the global geometry cannot be preserved if no additional side-information, e.g., anchor locations, is available [13]–[16]. To the best of our knowledge, unsupervised learning for localization remains an open challenge and has not been adequately addressed to date.

In this paper, as model-based and ML-based methods have their respective advantages, we aim to combine them to mutually enhance overall positioning performance. In summary, the main contributions of this work are given as follows: 1) A unified localization method is proposed, where LoS/NLoS identification is conducted, based on which the model-based method or a neural network (NN) model (obtained by unsupervised learning) is adopted for localization; 2) In order to reduce the reliance on exhaustive field measurement campaigns and ground-truth label collection, a ML-based method is presented in a manner of unsupervised learning, where self-generated labels are used for NN training. By resorting to optimal transport (OT) [18], [19], these self-generated training labels

[1]In this paper, NLoS scenarios refer to scenarios where LoS path is blocked. Conversely, LoS scenarios describe cases where the LoS path exists, regardless of whether NLoS paths are present or not.

are derived jointly from the position estimates of model-based methods and map information; 3) Ray-tracing simulations are carried out to evaluate the unified method, which was shown to outperform model-based methods and achieve a competitive performance with fingerprinting. We are basically bridging the performance gap (e.g., positioning accuracy) between unsupervised channel charting and fully-supervised fingerprinting, by leveraging minimal side-information (i.e., building layouts).

II. SYSTEM MODEL

We consider a single-cell orthogonal frequency division multiplexing (OFDM) system, where a base station (BS) equipped with M antennas serves multiple single-antenna users over N_c subcarriers. A dedicated sequence of pilots is transmitted by each user such that its associated CSI can be estimated by the BS. We assume perfect channel estimation at the BS, and the CSI from user i at all subcarriers is denoted by $\mathbf{H}_i \in \mathbb{C}^{M \times N_c}$. Moreover, tight synchronization between users and the BS is achievable through protocols like synchronization signal transmission, round-trip time (RTT) measurements, and timing advance mechanism [20]. Our goal is to estimate the position of each user, denoted by $\mathbf{p}_i = [x_i, y_i, z_i]^\mathsf{T}$, based solely on its associated CSI, i.e., $\hat{\mathbf{p}}_i = f_e(\mathbf{H}_i)$, where $\hat{\mathbf{p}}_i$ is the estimate of $\mathbf{p}_i$ and $f_e(\cdot)$ is an estimation function.

We assume that a 3D map of the environment is accessible, based on which the region for all possible user positions can be extracted and denoted by $\mathcal{R} \subseteq \mathbb{R}^{3 \times 1}$. Therefore, for any user position in the system, we have $\mathbf{p}_i \in \mathcal{R}, \forall i$. Moreover, the 3D map enables us to determine whether a given point in $\mathcal{R}$ has a LoS connection to the BS or not, based on which the regions for LoS and NLoS users can be distinguished. We denote the regions for LoS users and NLoS users as $\mathcal{R}_{\mathrm{LoS}} \subseteq \mathbb{R}^{3 \times 1}$ and $\mathcal{R}_{\mathrm{NLoS}} \subseteq \mathbb{R}^{3 \times 1}$, respectively, where $\mathcal{R}_{\mathrm{LoS}} \cap \mathcal{R}_{\mathrm{NLoS}} = \emptyset$ and $\mathcal{R}_{\mathrm{LoS}} \cup \mathcal{R}_{\mathrm{NLoS}} = \mathcal{R}$.

A uniform linear array (ULA) with inter-antenna separation of d is adopted at the BS. Considering multi-path effect, the frequency domain channel matrix for user i can be modeled as $\mathbf{H}_i = \sum_{l=1}^{L_i} \beta_l^i \mathbf{a_t}(\theta_l^i) \mathbf{b}^\mathsf{T}(\tau_l^i), \forall i$ [12], [21], where there are L_i channel paths for user i; β_l^i, θ_l^i, and τ_l^i are the complex channel gain, angle-of-arrival (AoA), and time-of-arrival (ToA) of the l-th path, respectively; $\mathbf{b}(\tau) \in \mathbb{C}^{N_c \times 1}$ is the frequency-domain steering vector, and $[\mathbf{b}(\tau)]_n = \exp[-\jmath 2\pi(n-1)\Delta_f \tau]$, $n = 1, 2, \cdots, N_c$, with Δ_f being the subcarrier spacing; $\mathbf{a_t}(\theta) \in \mathbb{C}^{M \times 1}$ is array steering vector at the BS, and $[\mathbf{a_t}(\theta)]_m = \exp[\jmath 2\pi(m-1)\frac{d}{\lambda}\sin(\theta)]$, $m = 1, 2, \cdots, M$, where $\lambda = c/f_c$ with c being the speed of light and f_c being the carrier frequency.

III. UNIFIED LOCALIZATION

The unified localization method is illustrated in Fig. 2. Firstly, LoS/NLoS identification is conducted based on CSI $\mathbf{H}_i$; then, if user i is identified as a LoS user, the model-based method would be used to estimate its position; otherwise, a NN model is adopted to generate $\hat{\mathbf{p}}_i$ based on $\mathbf{H}_i$. The model-based and NN-based methods are integrated as

$$\hat{\mathbf{p}}_i = \begin{cases} f_{\mathrm{pe}}\left[f_{\mathrm{ce}}(\mathbf{H}_i)\right], & \text{if } I'(\mathbf{H}_i) = 1; \\ f_{\Theta}\left[f_{\mathrm{extr}}(\mathbf{H}_i)\right], & \text{if } I'(\mathbf{H}_i) = 0, \end{cases} \quad (1)$$
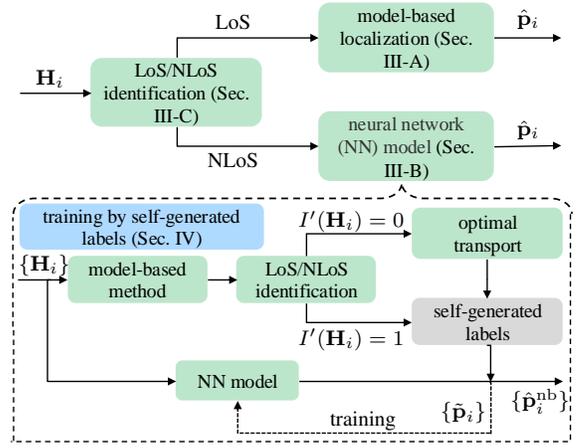
Figure 2: The unified localization method, where the dashed border represents the training process, including training label self-generation, of the NN model.

where $f_{\mathrm{ce}}(\cdot)$ is a channel parameter estimation function; $f_{\mathrm{pe}}(\cdot)$ is a function that maps channel parameters to position based on geometric relationships; $f_{\mathrm{extr}}(\cdot)$ is a channel feature extraction function to transform CSI $\mathbf{H}_i$ to the input of the NN model; $f_{\Theta}(\cdot)$ is the forward function of the NN model with Θ being parameters; and $I'(\cdot)$ is a CSI-based LoS/NLoS identification function. Note that the NN model is obtained by unsupervised learning, where self-generated training labels $\{\tilde{\mathbf{p}}_i\}$ are used to learn the relationship between CSI and user position. More details about each block are given as follows.

A. Model-based Method

With OFDM transmission and multi-antenna BS, channel parameters, including channel gain, AoA, and ToA for different channel paths, can be estimated by channel estimation algorithms, e.g., multiple signal classification (MUSIC) algorithm [9], based on which the position of a user can be derived by utilizing geometric relationships [4], [12], [21]. In this paper, the orthogonal matching pursuit (OMP) algorithm [10]–[12] is used, where $\hat{L}_i$ channel paths can be recovered for user i with channel parameters $\{\hat{\beta}_l^i, \hat{\theta}_l^i, \hat{\tau}_l^i\}$, $l = 1, 2, \ldots, \hat{L}_i$ (for more details, the reader is referred to Algorithm 1 of [12]). For a low complexity implementation, the shortest channel path among the recovered $\hat{L}_i$ paths is identified as the LoS path and used to estimate the user position. Thus, the channel parameter estimation function $f_{\mathrm{ce}}(\cdot)$ that maps the CSI to AoA and ToA can be written as $\{\hat{\theta}_s^i, \hat{\tau}_s^i\} = f_{\mathrm{ce}}(\mathbf{H}_i), \forall i$, where $s = \arg\min_l \hat{\tau}_l^i, l = 1, 2, \ldots, \hat{L}_i$.

With $\{\hat{\theta}_s^i, \hat{\tau}_s^i\}$ (yielding sufficient information to estimate x-coordinate and y-coordinate for a given z-coordinate), user position estimate can be derived by regarding this channel path as LoS, based on the following geometric relationships:

$$\hat{\tau}_s^i = \frac{\|\mathbf{p}_{\mathrm{BS}} - \hat{\mathbf{p}}_i\|_2}{c}, \quad \hat{\theta}_s^i = \arcsin\left(\frac{(\hat{\mathbf{p}}_i - \mathbf{p}_{\mathrm{BS}}) \cdot \mathbf{n}_{\mathrm{ULA}}}{\|\hat{\mathbf{p}}_i - \mathbf{p}_{\mathrm{BS}}\|_2}\right), \quad (2)$$

where $\mathbf{p}_{\mathrm{BS}}$ is the position of the BS and $\mathbf{n}_{\mathrm{ULA}}$ is the normalized vector representing the direction of the ULA. Therefore, the model-based estimate $\hat{\mathbf{p}}_i^{\mathrm{mb}}$ can be written as

$\hat{\mathbf{p}}_i^{\text{mb}} = f_{\text{pe}}[f_{\text{ce}}(\mathbf{H}_i)]$, $\forall i$, where $f_{\text{pe}}(\cdot)$ can be obtained, for example, by solving (2) (see, e.g., [1]–[5] for more details on geometric relationships).

B. Unsupervised Learning Method

In order to obtain a parametric model, NN is adopted to map CSI to position. Instead of directly feeding $\mathbf{H}_i$ to NN model, CSI preprocessing is used to extract features, which not only reduces the dimension of input data but also makes NN easier to learn hidden patterns [7], [8]. To do so, we first convert the antenna-frequency domain CSI $\mathbf{H}_i$ to the angle-delay domain, as given by $\bar{\mathbf{H}}_i = \mathcal{F}_a\left[\mathcal{F}_b^{-1}(\mathbf{H}_i)\right]$, $\forall i$, where $\mathcal{F}_a(\cdot)$ is the discrete Fourier transform along the antenna axis; $\mathcal{F}_b^{-1}(\cdot)$ is the inverse discrete Fourier transform along the frequency axis.

Since all channel paths in $\mathbf{H}_i$ have $\tau_l^i \in [\tau_{\min}, \tau_{\max}]$, where $\tau_{\min}$ and $\tau_{\max}$ are the minimum ToA and maximum ToA, the angle-delay domain CSI $\bar{\mathbf{H}}_i$ can be truncated by taking the columns that have delays within $\tau_{\min}$ and $\tau_{\max}$. Specifically, we truncate $\bar{\mathbf{H}}_i$ as $\bar{\mathbf{H}}_i^{\text{t}} = [\bar{\mathbf{H}}_i]_{:,\lfloor\tau_{\min}\Delta_f N_c\rfloor:\lceil\tau_{\max}\Delta_f N_c\rceil}$, $\forall i$, whose element-wise amplitude and phase can be used to generate the input to the NN[2], as denoted by $\mathbf{d}_i = \text{vec}\left(\left[\log(|\bar{\mathbf{H}}_i^{\text{t}}|), \angle\bar{\mathbf{H}}_i^{\text{t}}\right]\right)$, $\forall i$, where $\text{vec}(\cdot)$ would vectorize a matrix. Thus, implementing CSI domain conversion, truncation, and element-wise extraction, $f_{\text{extr}}(\cdot)$ transforms $\mathbf{H}_i$ to the input of the NN model, i.e., $\mathbf{d}_i = f_{\text{extr}}(\mathbf{H}_i)$, based on which the output of the NN model, i.e., the user position estimate $\hat{\mathbf{p}}_i^{\text{nb}}$, can be expressed by $\hat{\mathbf{p}}_i^{\text{nb}} = f_{\Theta}[f_{\text{extr}}(\mathbf{H}_i)]$, $\forall i$.

Given a training dataset $\mathcal{S} : \{\mathbf{H}_i\}_{i=1,2,...,N_u}$, where N_u is the number of user positions, an extended dataset $\mathcal{S}' : \{\mathbf{d}_i, \tilde{\mathbf{p}}_i\}_{i=1,2,...,N_u}$ can be obtained by applying $f_{\text{extr}}(\cdot)$ and training label self-generation. The self-generated labels $\{\tilde{\mathbf{p}}_i\}$ can be obtained based on CSI $\{\mathbf{H}_i\}$ and OT as shown in Fig. 2: for identified LoS users, $\{\tilde{\mathbf{p}}_i|I'(\mathbf{H}_i) = 1\}$ are the model-based estimates $\{\hat{\mathbf{p}}_i^{\text{mb}}|I'(\mathbf{H}_i) = 1\}$; while for identified NLoS users, $\{\tilde{\mathbf{p}}_i|I'(\mathbf{H}_i) = 0\}$ are obtained by transforming $\{\hat{\mathbf{p}}_i^{\text{mb}}|I'(\mathbf{H}_i) = 0\}$ via OT. More details are deferred to Sec. IV. Then, the loss function of the NN model can be designed as

$$\mathcal{L} = \frac{1}{N_u} \sum_{\mathbf{d}_i, \tilde{\mathbf{p}}_i \in \mathcal{S}'} \|f_{\Theta}(\mathbf{d}_i) - \tilde{\mathbf{p}}_i\|_2^2, \tag{3}$$

based on which the parameters Θ can be updated by back-propagation during training with the labels $\{\tilde{\mathbf{p}}_i\}$ that are self-generated before. As in Sec. V, when an L_{MLP}-layer multi-layer perceptron (MLP) with layer widths $\{n_l\}_{l=1,...,L_{\text{MLP}}}$ is adopted for the NN model, the complexity of training is $\mathcal{O}\left(N_u \sum_{l=1}^{L_{\text{MLP}}} n_{l-1}n_l\right)$ per epoch, where n_0 is the dimensionality of $\mathbf{d}_i$.

C. LoS/NLoS Identification

CSI-based LoS/NLoS identification has been widely studied and numerous algorithms are proposed, where some of them

achieve a high identification accuracy, e.g., 93%-100% for different system settings [4], [22]. It is assumed that we have a LoS/NLoS identification mechanism, denoted by $I(\cdot)$, where user i is identified as a LoS user if $I(\mathbf{H}_i) = 1$ and a NLoS user if $I(\mathbf{H}_i) = 0$. For the given $I(\cdot)$, the identification accuracy is denoted by $p_I \in [0.5, 1]$. Furthermore, with a LoS path, the mapping function $f_{\text{pe}}(\cdot)$, e.g., solving (2) based on AoA and ToA, would result in a position in the LoS region. This allows us to conclude that all LoS users would be localized in the LoS region with $f_{\text{pe}}[f_{\text{ce}}(\cdot)]$, i.e., if a user is localized in the NLoS region, it must be a NLoS user[3]. Therefore, given the results from $I(\mathbf{H}_i)$, we can further improve the LoS/NLoS identification by

$$I'(\mathbf{H}_i) = \begin{cases} 0, & \text{if } f_{\text{pe}}[f_{\text{ce}}(\mathbf{H}_i)] \in \mathcal{R}_{\text{NLoS}}; \\ I(\mathbf{H}_i), & \text{if } f_{\text{pe}}[f_{\text{ce}}(\mathbf{H}_i)] \in \mathcal{R}_{\text{LoS}}. \end{cases} \tag{4}$$

Remark. *When the mechanism $I(\cdot)$ is not accessible or its identification accuracy p_I drops to a extremely low value, a conservative method that does not rely on $I(\cdot)$ can be presented. In particular, without $I(\cdot)$, the model-based results could be used for LoS/NLoS identification, i.e., if a user is localized in the LoS (resp. NLoS) region based on $f_{\text{pe}}[f_{\text{ce}}(\cdot)]$, it would be identified as a LoS (resp. NLoS) user. Therefore, the LoS/NLoS identification (4) can be reformulated as*

$$I'(\mathbf{H}_i) = \begin{cases} 0, & \text{if } f_{\text{pe}}[f_{\text{ce}}(\mathbf{H}_i)] \in \mathcal{R}_{\text{NLoS}}; \\ 1, & \text{if } f_{\text{pe}}[f_{\text{ce}}(\mathbf{H}_i)] \in \mathcal{R}_{\text{LoS}}, \end{cases} \tag{5}$$

based on which LoS and NLoS users can be identified completely by $f_{\text{pe}}[f_{\text{ce}}(\mathbf{H}_i)]$. Then, referring to (1), the position estimates for identified LoS (resp. NLoS) users are obtained by $\hat{\mathbf{p}}_i = f_{\text{pe}}[f_{\text{ce}}(\mathbf{H}_i)]$ (resp. $\hat{\mathbf{p}}_i = f_{\Theta}[f_{\text{extr}}(\mathbf{H}_i)]$).

IV. OT-BASED LABEL SELF-GENERATION

A. Discrete Optimal Transport

Given a source domain $\mathcal{X}_s$ with a defined probability measure u_s and a target domain $\mathcal{X}_t$ with a defined probability measure u_t, OT aims to find a transformation from $\mathcal{X}_s$ to $\mathcal{X}_t$, as denoted by $\mathbf{T}^*(\cdot)$, that minimizes the transportation cost on the condition that the image measure of u_s in $\mathcal{X}_t$, as denoted by u_s' (since $\mathbf{T}^*(\cdot)$ transforms u_s to u_s', we have $u_s(\mathbf{x}_s) = u_s'[\mathbf{T}^*(\mathbf{x}_s)]$, $\forall \mathbf{x}_s \in \mathcal{X}_s$), is $u_s'(\mathbf{x}_t) = u_t(\mathbf{x}_t)$, $\forall \mathbf{x}_t \in \mathcal{X}_t$. Therefore, the OT problem can be written as [18]

$$\mathbf{T}^* = \arg\min_{\mathbf{T}} \int_{\mathcal{X}_s} c[\mathbf{x}_s, \mathbf{T}(\mathbf{x}_s)]\mathrm{d}u_s(\mathbf{x}_s), \tag{6}$$
$$\text{s.t.} \quad u_s(\mathbf{x}_s) = u_t[\mathbf{T}(\mathbf{x}_s)], \forall \mathbf{x}_s \in \mathcal{X}_s,$$

where $c[\mathbf{x}_s, \mathbf{T}(\mathbf{x}_s)]$ is the transportation cost function to transform $\mathbf{x}_s$ to $\mathbf{T}(\mathbf{x}_s)$.

The convex relaxation of the above problem (6) can be formulated as finding a joint probability function γ with marginals u_s and u_t to minimize $\int_{\mathcal{X}_s \times \mathcal{X}_t} c(\mathbf{x}_s, \mathbf{x}_t)d\gamma(\mathbf{x}_s, \mathbf{x}_t)$ [19]. Moreover, when $\mathcal{X}_s = \{\mathbf{x}_s^i\}_{i=1,2,...,N_s}$ and $\mathcal{X}_t = \{\mathbf{x}_t^i\}_{i=1,2,...,N_t}$, where N_s and N_t are the number of data samples in $\mathcal{X}_s$ and $\mathcal{X}_t$, the probability measures u_s and u_t can be given in the form of $\mathbf{u}_s$ and $\mathbf{u}_t$, which are the probability vectors of the data samples in $\mathcal{X}_s$ and $\mathcal{X}_t$, respectively. Therefore, optimizing the

joint probability matrix $\mathbf{\Gamma} \in \mathbb{R}_+^{N_s \times N_t}$ across $\mathcal{X}_s$ and $\mathcal{X}_t$, the problem (6) in discrete case can be approximated as [19]

$$\mathbf{\Gamma}^* = \arg\min_{\mathbf{\Gamma} \in \mathcal{B}} \quad \langle \mathbf{\Gamma}, \mathbf{C} \rangle_F, \tag{7}$$

where $\langle \cdot, \cdot \rangle_F$ is the Frobenius dot product, $\mathbf{C} \in \mathbb{R}_+^{N_s \times N_t}$ is the transportation cost matrix with $[\mathbf{C}]_{i,j} = c(\mathbf{x}_s^i, \mathbf{x}_t^j)$, and

$$\mathcal{B} = \{ \mathbf{\Gamma} \in \mathbb{R}_+^{N_s \times N_t} | \mathbf{\Gamma} \cdot \mathbf{1}_{N_s} = \mathbf{u}_s, \ \mathbf{\Gamma}^\mathsf{T} \cdot \mathbf{1}_{N_t} = \mathbf{u}_t \}, \tag{8}$$

with $\mathbf{1}_N$ being a all-one vector with the size of N. Note that $\mathcal{B}$ represents all possible joint probability matrices $\{\mathbf{\Gamma}\}$ that are with marginals $\mathbf{u}_s$ and $\mathbf{u}_t$. It is observed that the discrete OT problem (7) is a linear programming and can be solved efficiently by, e.g., simplex methods [18], which usually infer a polynomial time complexity. Moreover, low-complexity algorithms can be adopted to approximately solve (7), e.g., the Sinkhorn algorithm [18] with a complexity of $\mathcal{O}\left(I_{\mathrm{Iter}} N_s N_t\right)$, where I_{Iter} is the number of Sinkhorn iterations.

B. Self-generated Training Labels

Given the training dataset $\mathcal{S}: \{\mathbf{H}_i\}_{i=1,2,\dots,N_u}$, position estimates $\{\hat{\mathbf{p}}_i^{\mathrm{mb}}\}_{i=1,2,\dots,N_u}$ can be attained based on the model-based method $\hat{\mathbf{p}}_i^{\mathrm{mb}} = f_{\mathrm{pe}}\left[f_{\mathrm{ce}}(\mathbf{H}_i)\right]$. For identified LoS users, $\{\hat{\mathbf{p}}_i^{\mathrm{mb}}\}$ can be used as training labels, i.e., $\tilde{\mathbf{p}}_i = \hat{\mathbf{p}}_i^{\mathrm{mb}}$ if $I'(\mathbf{H}_i) = 1$; however, these estimates $\hat{\mathbf{p}}_i^{\mathrm{mb}}$ may not be accurate for identified NLoS users and do not match the map. Therefore, the next task is to find a transformation $\mathbf{T}^*(\cdot)$ that maps $\{\hat{\mathbf{p}}_i^{\mathrm{mb}} | I'(\mathbf{H}_i) = 0\}_{i=1,2,\dots,N_u}$ to some other positions that can match the NLoS region on the map.

To achieve this goal, we resort to the discrete OT (7), where the source domain $\mathcal{X}_s = \{\hat{\mathbf{p}}_i^{\mathrm{mb}} | I'(\mathbf{H}_i) = 0\}_{i=1,2,\dots,N_u}$ and the target domain $\mathcal{X}_t = \{\bar{\mathbf{p}}_i\}_{i=1,2,\dots,N_t}$, based on which the joint probability matrix $\mathbf{\Gamma} \in \mathbb{R}_+^{N_s \times N_t}$ is optimized ($N_s = N_u - \sum_i I'(\mathbf{H}_i)$). In this paper, the cost associated with two positions is set as the squared Euclidean distance[4], i.e., $[\mathbf{C}]_{i',j} = \|\hat{\mathbf{p}}_i^{\mathrm{mb}} - \bar{\mathbf{p}}_j\|_2^2$. Based on the map, the data samples in the target domain $\{\bar{\mathbf{p}}_i\}_{i=1,2,\dots,N_t}$ can be generated by selecting N_t grid points in the NLoS region, i.e., $\bar{\mathbf{p}}_i \in \mathcal{G}_{\mathrm{NLoS}} \subseteq \mathcal{R}_{\mathrm{NLoS}}$, where $\mathcal{G}_{\mathrm{NLoS}}$ contains all grid points with a spacing distance of Δ_d in $\mathcal{R}_{\mathrm{NLoS}}$. If all users are located uniformly on the map[5], the probability vectors $\mathbf{u}_s$ and $\mathbf{u}_t$ can be written as

$$\mathbf{u}_s = \frac{1}{N_s} \cdot \mathbf{1}_{N_s}, \ \mathbf{u}_t = \frac{1}{N_t} \cdot \mathbf{1}_{N_t}. \tag{9}$$

Once the optimal joint probability matrix $\mathbf{\Gamma}^*$ is obtained, the transformed position for each identified NLoS user can be expressed by [18]

$$\tilde{\mathbf{P}}_{\mathrm{NLoS}} = \mathrm{diag}(\mathbf{\Gamma}^* \cdot \mathbf{1}_{N_t})^{-1} \cdot \mathbf{\Gamma}^* \cdot \bar{\mathbf{P}}, \tag{10}$$

where $\bar{\mathbf{P}} \in \mathbb{R}^{N_t \times 3}$ is the collection of all $\{\bar{\mathbf{p}}_i\}_{i=1,2,\dots,N_t}$; $\tilde{\mathbf{P}}_{\mathrm{NLoS}} \in \mathbb{R}^{N_s \times 3}$ with $[\tilde{\mathbf{P}}_{\mathrm{NLoS}}]_{i',:} = \mathbf{T}^*(\hat{\mathbf{p}}_i^{\mathrm{mb}})$ being the mapped position[4]. Then, combining the identified LoS and NLoS users, the self-generated labels $\{\tilde{\mathbf{p}}_i\}_{i=1,2,\dots,N_u}$ can expressed as

$$\tilde{\mathbf{p}}_i = \begin{cases} \hat{\mathbf{p}}_i^{\mathrm{mb}}, & \text{if } I'(\mathbf{H}_i) = 1; \\ \mathbf{T}^*(\hat{\mathbf{p}}_i^{\mathrm{mb}}), & \text{if } I'(\mathbf{H}_i) = 0. \end{cases} \tag{11}$$

---

[4]Note that the i'-th data sample in $\{\hat{\mathbf{p}}_i^{\mathrm{mb}} | I'(\mathbf{H}_i) = 0\}_{i=1,2,\dots,N_u}$ is the i-th data sample in $\{\hat{\mathbf{p}}_i^{\mathrm{mb}}\}_{i=1,2,\dots,N_u}$.

[5]Note that the assumption of uniformly distributed users on the map is not mandatory; on the contrary, any distribution is allowed in general for OT.

V. NUMERICAL EXPERIMENTS

A. System Setting

In this section, numerical experiments in a street canyon scenario, as shown in Fig. 1, are carried out to evaluate the unified localization method. The BS is placed at $\mathbf{p}_{\mathrm{BS}} = [0, -9, 57]^\mathsf{T}$ with $M = 256$ antennas (the inter-antenna distance is set as $d = \lambda/2$), and the height of each user is set as $1.5\,\mathrm{m}$, i.e., $\mathbf{p}_i = [x_i, y_i, 1.5]^\mathsf{T}$. The carrier frequency of the system is $10\,\mathrm{GHz}$, and $50\,\mathrm{MHz}$ bandwidth is used with subcarrier spacing of $120\,\mathrm{kHz}$. For a better illustration, users are placed randomly in the first quadrant of the map, i.e., $\mathbf{p}_i \in \mathcal{R}$ and $\mathbf{p}_i \in \mathbb{R}_+^{3 \times 1}$, $\forall i$. For each user, the CSI $\mathbf{H}_i$ is generated realistically by Sionna library [17], where ray-tracing techniques are used to generate the propagation paths, based on which the channel coefficients are computed. When using OT, N_t is set as the total number of all grid points in $\mathcal{G}_{\mathrm{NLoS}}$ with $\Delta_d = 0.5\,\mathrm{m}$.

As for the ML-based methods, following the same NN architecture and hyperparameters as in [15], [16], a MLP is adopted for $f_\Theta(\cdot)$, which consists of 5 hidden layers (each with 1024, 512, 256, 128, 64 neurons, respectively, ReLU activation, and batch normalization) and an output layer (with 2 neurons and linear activation). Around 1800 user positions, consisting of 550 LoS and 1250 NLoS, are generated both for training and testing datasets. Moreover, Adam optimizer with a decayed learning rate is employed for training (1600 epochs).

B. Benchmarks

In this section, mean absolute error (MAE) [6], [14], [15], i.e., $\frac{1}{N_u} \sum_i \|\hat{\mathbf{p}}_i - \mathbf{p}_i\|_2$, is adopted as a metric for evaluation and analysis. For comparison, in addition to the unified method (1) and its conservative variant, model-based, fingerprinting, and channel charting methods are considered as benchmarks:

- **Model-based method**: position estimates are obtained based on the OMP algorithm and geometric relationships, i.e., $\hat{\mathbf{p}}_i^{\mathrm{mb}} = f_{\mathrm{pe}}\left[f_{\mathrm{ce}}(\mathbf{H}_i)\right]$, $\forall i$;
- **Fingerprinting**: position estimates are obtained from NN, i.e., $\hat{\mathbf{p}}_i^{\mathrm{nb}} = f_\Theta\left[f_{\mathrm{extr}}(\mathbf{H}_i)\right]$, where $f_\Theta(\cdot)$ is trained by assuming ground-truth labels, i.e., $\tilde{\mathbf{p}}_i = \mathbf{p}_i$, $\forall i$. It is a fully-supervised learning approach;
- **Channel charting**[6]: position estimates are obtained from NN, i.e., $\hat{\mathbf{p}}_i^{\mathrm{nb}} = f_\Theta\left[f_{\mathrm{extr}}(\mathbf{H}_i)\right]$, where $f_\Theta(\cdot)$ is trained based on the following loss function: $\mathcal{L}_{\mathrm{CC}} = \frac{1}{N_u^2} \sum_{\mathbf{H}_i \in \mathcal{S}} \sum_{\mathbf{H}_j \in \mathcal{S}} \frac{(\|f_\Theta(\mathbf{d}_i) - f_\Theta(\mathbf{d}_j)\|_2 - \varepsilon d_{i,j})^2}{d_{i,j}} + \frac{1}{\sum_k I'(\mathbf{H}_k)} \sum_{\{\mathbf{H}_k | I'(\mathbf{H}_k)=1\}} \|f_\Theta(\mathbf{d}_k) - \hat{\mathbf{p}}_k^{\mathrm{mb}}\|_2^2$, where ε is a scaling parameter (learned by training), and $d_{i,j}$ is the feature "distance" between two CSIs that would be preserved. In this paper, cosine dissimilarity is used for $\{d_{i,j}\}$ [15]. It is a fully unsupervised learning approach.

---

[6]Channel charting is, in general, a self-supervised learning technique that aims to generate a mapping from high-dimensional CSI to a low-dimensional space, which is called as channel chart, where nearby points correspond to nearby locations in real space. More details are referred to [13]–[16]. In this paper, since precise position estimates can be obtained by model-based methods for LoS users, a loss term corresponding to the estimation accuracy of the LoS users is added.
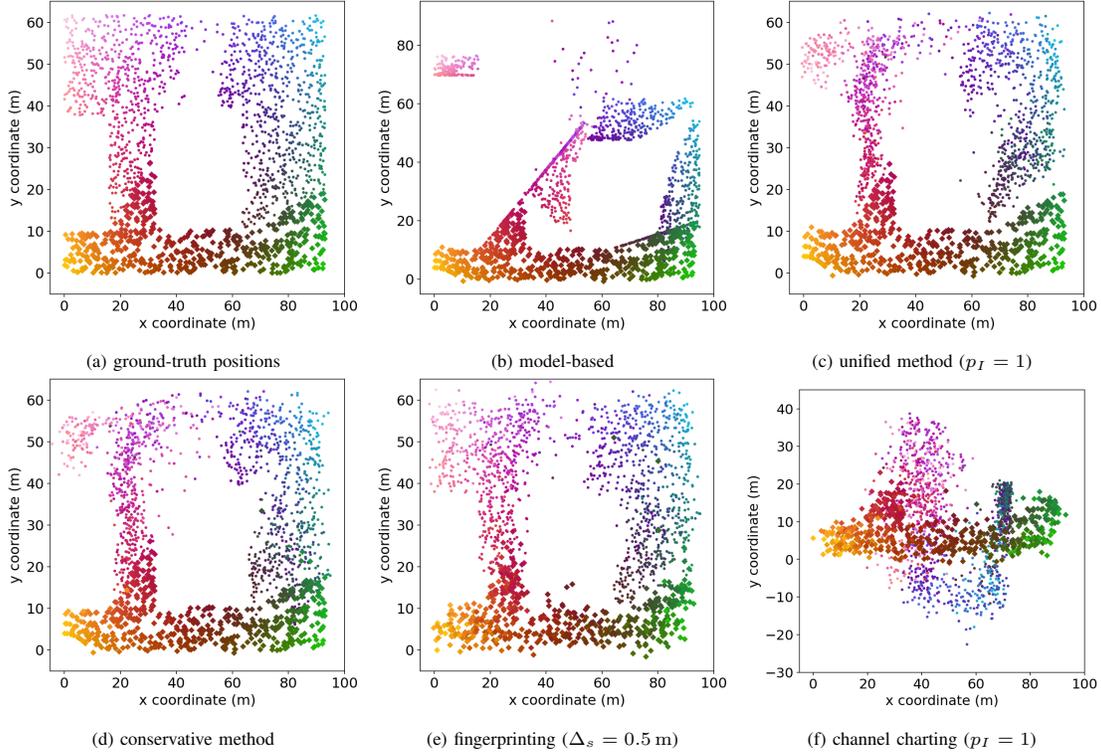
Figure 3: Position estimates for different localization methods (testing dataset; every user is associated with a unique color; dot marks represent NLoS users and square marks represent LoS users).

For a fair comparison, as fingerprinting requires extensive CSI measurement and collection, the dataset often contains limited data samples. For fingerprinting, grid positions on the map with corresponding CSI are used for training. Specifically, around 1800 user positions are selected from $\mathcal{G}$, which contains all grid points with a spacing distance of Δ_s on the map, i.e., $\mathbf{p}_i \in \mathcal{G} \subseteq \mathcal{R}$, $\forall i$. Note that if $\mathcal{G}$ has less than 1800 positions, all grid positions in $\mathcal{G}$ are selected. While for testing, user positions are placed randomly, which is the same as other considered methods.

C. Numerical Results

1) Comparison with benchmarks

The qualitative results (position estimates) obtained from different localization methods for the testing dataset are shown in Fig. 3. It is observed that the model-based method achieves a high positioning accuracy for LoS users (with a MAE of 0.34 m), while the estimates of NLoS users are distorted significantly. The fingerprinting method keeps the global shape of NLoS users, while the estimates of LoS users are not as precise as the model-based method. Being an integrated strategy, the unified method (also the conservative variant) can not only attain a precise estimation for LoS users, but also the global position of the NLoS users can be preserved, which shows the effectiveness of the proposed OT-based approach in Sec. IV in mapping the NLoS users to feasible locations that respect the map constraints. Moreover, although the local geometry, i.e., the neighboring relationship between users, is preserved, channel charting cannot capture the global geometry of NLoS
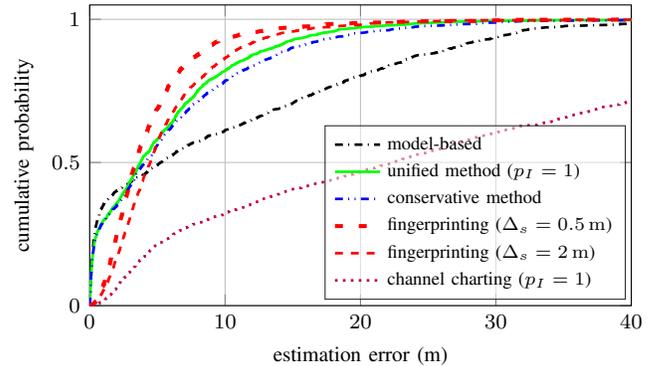

Figure 4: The cumulative distribution function (CDF) w.r.t. the positioning error for different methods (testing dataset).

users due to the lack of global coordinate information, resulting in a high estimation error for NLoS users. This indicates that the model-based methods could provide important information for unsupervised learning methods to improve the estimation accuracy, a strategy followed by the proposed unified method.

For a quantitative evaluation of the performance, the CDF and MAE for all considered methods are shown in Fig. 4 and Table I, respectively. It is seen that the model-based method and the unified method (with $p_I = 1$) achieve the lowest estimation error for LoS users; while the fingerprinting with a small Δ_s has the lowest estimation error for NLoS users and attains the highest overall performance for all users. The unified method significantly outperforms the model-based method, and achieves an estimation accuracy comparable to that of

Table I: MAE for different methods (testing dataset).

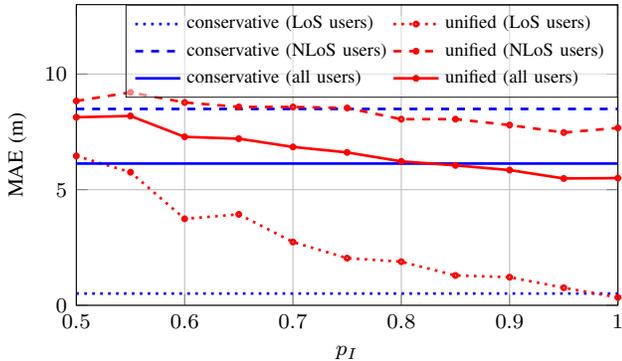| MAE (m)	LoS	NLoS	all users
Unsupervised Methods			
model-based	**0.34**	13.79	9.81
unified method ($p_I = 1$)	**0.34**	**7.67**	**5.50**
conservative method	0.51	8.49	6.13
channel charting ($p_I = 1$)	4.19	36.02	26.60
Supervised Methods			
fingerprinting ($\Delta_s = 2\,\mathrm{m}$)	4.57	6.13	5.67
fingerprinting ($\Delta_s = 1\,\mathrm{m}$)	**3.94**	5.43	4.99
fingerprinting ($\Delta_s = 0.5\,\mathrm{m}$)	3.98	**4.77**	**4.54**

Figure 5: MAE for the unified localization method with different p_I (testing dataset).

fingerprinting with $\Delta_s = 0.5\,\mathrm{m}$ (within 1 m in terms of MAE). Notably, with unsupervised learning, the unified method is able to achieve a slightly better overall performance than the fully-supervised fingerprinting with $\Delta_s = 2\,\mathrm{m}$.

2) Discussion on p_I

In order to explore the impact of the LoS/NLoS identification accuracy, Fig. 5 shows the MAE for the unified method with different p_I and its conservative variant. It is clear (also intuitive) that the performance of the unified method relies on the identification accuracy p_I, and its estimation error would increase with the reduction of p_I. Moreover, the performance for LoS users degrades much faster, while the NLoS users are more robust to the identification error. This is because the wrong identification for some NLoS users (identified as LoS users by $I(\mathbf{H}_i)$ but with $f_{\mathrm{pe}}\left[f_{\mathrm{ce}}(\mathbf{H}_i)\right] \in \mathcal{R}_{\mathrm{NLoS}}$) can be corrected based on the map information, as shown in (4).

It also can be seen that for a reasonable $p_I \in [0.93, 1]$ (from the existing identification methods in the literature), the unified method performs better than its conservative variant overall. There is a critical $p_I \approx 0.83$, corresponding to the intersection point of the overall MAE, below which the conservative method is a better choice. This implies that a dynamic switch between the unified and conservative methods based on p_I can be adopted: when p_I is larger to some threshold, e.g., 0.83, the unified method is used; otherwise, the conservative method is used.

VI. CONCLUSIONS

In this paper, we proposed a unified localization method, along with a conservative variant, that reaps the benefits of model-based methods for LoS users and ML-based methods for NLoS users. In this method, OT was adopted to self-generate labels such that NN can be trained in a fully unsupervised manner without the prohibitive cost of data measurement and collection. Numerical results showed that the proposed method delivers competitive overall performance with the fully-supervised fingerprinting. Moreover, a threshold w.r.t. the LoS/NLoS identification accuracy was discovered such that an appropriate choice between the unified method and its conservative variant can be made for better performance.

REFERENCES

[1] H. Chen, H. Sarieddeen, T. Ballal, H. Wymeersch *et al.*, "A tutorial on terahertz-band localization for 6G communication systems," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 3, pp. 1780–1815, 2022.

[2] A. Shahmansoori, G. E. Garcia, G. Destino *et al.*, "Position and orientation estimation through millimeter-wave MIMO in 5G systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1822–1835, 2018.

[3] R. Mendrzik *et al.*, "Harnessing NLOS components for position and orientation estimation in 5G millimeter wave MIMO," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 93–107, 2019.

[4] R. Zekavat and R. M. Buehrer, *An Introduction to NLOS Identification and Localization*, 2012, pp. 523–555.

[5] O. Kaltiokallio, E. Rastorgueva-Foi *et al.*, "Robust snapshot radio SLAM," *IEEE Trans. Veh. Technol.*, vol. 74, no. 5, pp. 8460–8465, 2025.

[6] N. Singh, S. Choe, and R. Punmiya, "Machine learning based indoor localization using Wi-Fi RSSI fingerprints: An overview," *IEEE Access*, vol. 9, pp. 127 150–127 174, 2021.

[7] Y. Chen, N. González-Prelcic, T. Shimizu, and H. Lu, "Learning to localize with attention: from sparse mmWave channel estimates from a single BS to high accuracy 3D location," 2025.

[8] X. Gong *et al.*, "Machine learning-based fingerprint positioning for massive MIMO systems," *IEEE Access*, vol. 10, pp. 89 320–89 330, 2022.

[9] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

[10] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[11] J. Lee *et al.*, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, 2016.

[12] J. Miguel Mateos-Ramos, C. Häger, M. F. Keskin *et al.*, "Model-based end-to-end learning for multi-target integrated sensing and communication under hardware impairments," *IEEE Trans. Wireless Commun.*, vol. 24, no. 3, pp. 2574–2589, 2025.

[13] C. Studer *et al.*, "Channel charting: Locating users within the radio environment using channel state information," *IEEE Access*, vol. 6, pp. 47 682–47 698, 2018.

[14] S. Taner *et al.*, "Channel charting in real-world coordinates with distributed MIMO," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2025.

[15] P. Stephan, F. Euchner, and S. T. Brink, "Angle-delay profile-based and timestamp-aided dissimilarity metrics for channel charting," *IEEE Trans. Commun.*, vol. 72, no. 9, pp. 5611–5625, 2024.

[16] P. Ferrand, A. Decurninge, L. G. Ordoñez, and M. Guillaud, "Triplet-based wireless channel charting: Architecture and experiments," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2361–2373, 2021.

[17] J. Hoydis, S. Cammerer, F. Ait Aoudia, M. Nimier-David, L. Maggi, G. Marcus *et al.*, "Sionna," 2022, https://nvlabs.github.io/sionna/.

[18] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy, "Optimal transport for domain adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 9, pp. 1853–1865, 2017.

[19] G. Peyré and M. Cuturi, "Computational optimal transport: With applications to data science," *Found. Trends Mach. Learn.*, vol. 11, no. 5-6, pp. 355–607, 2019.

[20] 3rd Generation Partnership Project (3GPP), "Physical layer procedures for control," 3GPP, TS 38.213, Mar. 2025.

[21] V. Koivunen, M. F. Keskin *et al.*, "Multicarrier ISAC: Advances in waveform design, signal processing, and learning under nonidealities," *IEEE Signal Process. Mag.*, vol. 41, no. 5, pp. 17–30, 2024.

[22] J. Schroeder, S. Galler *et al.*, "NLOS detection algorithms for ultra-wideband localization," in *Proc. WPNC*, 2007, pp. 159–166.